Guitar String Classification using Inharmonicity Trajectories for Automatic Tablature Transcription

Jonathan Michelson

June 2017

School of Music
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Dr. Richard Stern, Dr. Thomas Sullivan

*Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Music and Technology*

June 23, 2017
DRAFT

Abstract

We propose and evaluate two additional inharmonicity-based guitar string classification methods using only audio. We characterize each guitar string as the linear trajectory of its log-inharmonicity with respect to its fundamental pitch, then assign unseen notes to the strings whose trajectories best approximate the notes' estimated inharmonicities. Results from classical, acoustic, and electric guitars in the RWC instruments dataset show that this method performs slightly better than the existing inharmonicity-based method for test guitars on which the system wasn't exclusively trained. We also derive and evaluate an inharmonicity compensation feature to allow for tablature transcription of arbitrarily-tuned guitars, provided the tuning is known. Performance increases on personally recorded alternately-tuned guitars validate this tuning adjustment feature.

Acknowledgments

My sincere gratitude goes to Rich Stern for advising me every step of this journey. I'd never before tackled a research project of this scope, but Rich's guidance, patience, and insight somehow made this thesis happen. It's been an immeasurably rewarding experience, and I owe that to him. I'll miss our weekly one-on-one meetings that evolved into friendly chats about music, family, and life.

I'm indebted to Tom Sullivan for lending his qualified examination of this thesis. I'm also indebted to him in the literal sense as he welcomed me to Pittsburgh by treating me to lunch. His hospitality and accessibility helped me know I was in the right place.

I'd like to thank Matthieu Hodgkinson for sharing with me his MAT algorithm code, as well as Jacob Abesser and Isabel Barbancho for their helpful correspondence.

My peers Che-yuan Liang and Raymond Xia have both, of their own accord, contributed meaningfully to this work through our many brainstorming and discussion sessions, and for them I'm grateful. Many thanks to Nick Pourazima, whose Fender Telecaster we recorded for experiments.

I need to thank the remainder of my CMU graduate music technology cohort – Devansh Zurale, Steven Krenn, Chen Liang, Mutian Fu, Hao Huang, Anirudh Mani, Garrett Osborn, and Nick Pourazima – for the support role they played by simply being the fun constructive community that they were.

Finally, I acknowledge my mom and dad, to whom all success is due. I'd also like to thank my brother Nick Michelson solely for persuading me to abandon writing a grandiose Acknowledgments section, and definitely nothing else.

Contents

List of Figures

List of Tables

Chapter 1:   Introduction

The guitar is a popular musical instrument whose family comprises a diverse collection of stringed instruments. Its most prominent classes are the acoustic, classical, and electric guitars, and intraclass variation is high in characteristics such as body shape, string material, and amplification modality. Despite this diversity, the majority of guitar configurations exhibit some unifying features, like having six strings tuned to $E2$ (82Hz), $A2$ (110Hz), $D3$ (147Hz), $G3$ (196Hz), $B3$ (247Hz), and $E4$ (330Hz), for example. Performance is relatively constant as well; to play, a performer presses these strings with one hand against the neck's metal frets, each of which increase the string's pitch by one half-step, while strumming or plucking them with the other hand. Guitars typically have upwards of 20 frets, allowing for versatile musical passage realizations along the fretboard.

A consequence of the strings' open-tunings and extensive fretting locations is liberal pitch overlap between neighboring strings; multiple fretboard locations can usually be selected to produce a given pitch. Conventional music scores that represent passages as notes and chords therefore fail to communicate fretboard position. This ambiguity renders scores unfit for guitar students trying to learn more skillful placement of scales and arpeggios, or for enthusiasts trying to decrypt the fretboard positions used on recordings of virtuosic players' riffs. Figure 1.1 illustrates this ambiguity.

Tablature, on the other hand, is an alternative music notation that doesn't suffer from the one-to-many mapping of scores. The staff, instead of representing pitch as in classical scores, depicts a birds-eye view of the guitar neck from the performer's perspective. Each of the six horizontal

lines signify a corresponding string on the guitar, and numbers on each line specify the fret to be

played. Time progresses from left to right as in scores.
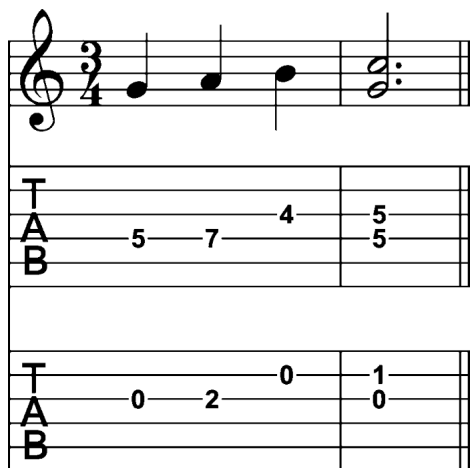
Figure 1.1: Musical score and two identical realizations of that score in tablature notation. Figure taken from [4].

Tablature (abbr. tabs) is widely more popular than scores for a number of likely reasons [21]. Its accessibility is attractive to those without formal musical training; interpreting a pictorial representation of the fretboard poses a smaller learning curve than studying music theory and deciphering score notation. Additionally, the lines and numbers in tablature that represent the strings and frets are easily represented with ASCII characters, rendering it an exceptionally portable mode of communication. Indeed, one can create, modify, and transfer tablature files using simply a text editor.

Accurate tablature transcription is a fairly laborious task, however. Dependable tabs containing few errors typically require a seasoned musician's careful listening to an audio recording, and possibly video recordings for additional consultation too. Reliable automation of this transcription task would benefit the guitar student community by expediting quality tab generation and unearthing fretboard positions of guitar riffs in songs for which accurate tabs don't yet exist. Recent technologies proposed to encourage proliferation of accessible tab transcription systems, such as web framework Robotaba [6], demonstrate a desire among MIR researchers to see this theoretically mature but practically nascent field mature into student-benefitting applications.

An important feature in systems that transcribe tablature from audio input is inharmonicity. Briefly, inharmonicity is a phenomenon that skews a vibrating string's harmonic partials upwards in frequency and whose variation trend is somewhat unique per string. Because of this, it is highly discriminative and has successfully been used as the sole feature in a transcription system by a team whose work [4] we consider the benchmark inharmonicity-based method. This is discussed further in Chapter 2, but for now its introduced simply for context.

This work builds on the benchmark method. We propose an alternative procedure to exploit the inharmonicity of an unseen note for string classification: we assign to it the string whose pre-computed log-inharmonicity trajectory predicts the note's estimated log-inharmonicity with minimal error. We evaluate the extent to which our performance differs from that of the benchmark system, focusing specifically on two cases: when the systems have been trained on recordings of only the test guitar, and when the systems have been trained on recordings of multiple guitars including the test guitar (all of which were in the same classical, acoustic, or electric class). Results show that although the benchmark system excels in the former case, ours succeeds in the latter, suggesting our method's greater generalizability. Additionally, this work introduces an inharmonicity compensation factor that allows for successful transcription of guitars whose tunings differ from those of the training set, provided that the alternate-tunings are known. Experiments on two personally recorded guitars in various tunings reveal that our compensation factor can restore an arbitrarily-tuned guitar's transcription performance to that of its more accurately transcribed standard-tuned trial.

In the next chapter, we survey previous work in the field of automatic tablature transcription, highlighting the benchmark inharmonicity-based method. This is followed by a more substantial motivation of inharmonicity's suitability for the string discrimination task and a review of the literature on inharmonicity estimation. In Chapter 3, we introduce our system in detail. Chapter 4 reports the results of our experiments, and Chapter 5 contains analysis of the results, threats to validity, and future work. We summarize this work in Chapter 6.

Chapter 2:   Previous Work

Music transcription is a rich research topic with many contributors. We begin with a brief review of various teams who have focused specifically on tablature transcription. We then introduce a feature of guitar audio referred to as inharmonicity, which has been used exclusively and successfully in one system, whose detailed summary we defer to Chapter 3. We close this chapter with a survey of the inharmonicity estimation literature.

2.1   Automatic Tablature Transcription

2.1.1   Graph Search Methods

A popular framework for the stringed-instrument transcription task, introduced by Sayegh [27] for both transcription and fingering problems, is that of weighted directed graphs that capture all plausible sequences of fretboard positions for a given musical passage. Edge weights in the graph are typically informed by mechanical or musical desirability of the fretboard sequences implied by the vertex transitions. Radicioni [24] used this graph paradigm to obtain optimal fingerings from music scores. Yazawa [31] extracted pitch and melody information from guitar recordings, and computed an optimal fingering based on plausible multipitch estimation results. Burlet [6] developed a polyphonic extension to monophonic optimal search algorithms. Burlet [7] also introduced a guitar-specific deep belief network whose polyphonic transcription output was arranged into tablature with a graph search routine. Radisavljevic [25] proposed a method to learn the optimal weights on the graph given optimal tablature.

2.1.2   Machine Learning Solutions

A number of teams have used machine learning techniques to ascertain fretboard position or related parameters. Gagnon [12] proposed using neural networks on scores to deduce general fretboard region position and number of strings playing, while Tuohy [30] used a genetic algorithm on scores to obtain tablature. Barbancho [3] attempted guitar string classification with a Fisher linear discriminant using a multitude of time and frequency-domain features. Barbancho [2] used HMMs on polyphonic recordings of common guitar chord sequences to transcribe their fret position and fingering. Abesser [1], Dittmar [9], and Kehling [19] achieve good string transcription performance after training an SVM on standard-tuning guitar audio and using a slew of partials-related features.

2.1.3   Multimedia Approaches

Some researchers have explored usage of media other than audio or musical information (i.e., scores). O'grady [22] exploited hexaphonic pickups and matrix factorization of their output to record fretboard positions. Paleari [23], Burns [8], and Kerdvibulvich [20] extract tablature from video recordings using image processing and computer vision techniques.

2.1.4   The Partial Coincidence Tally (PCT) Method

One system in particular served as inspiration for this work. Barbancho [4] realized the discriminative potential of a singular feature known as inharmonicity, and devised an algorithm that exclusively used inharmonicity to achieve impressive transcription results using only audio. We discuss inharmonicity in detail and methods for its estimation in the remainder of this chapter. Their method, which we refer to as the partial coincidence tally (PCT) method, is also given a more thorough treatment later in Chapter 3.

We were particularly interested in the PCT method for a number of reasons. Firstly, its in-

put is audio. Among the most interesting use cases for an automatic tab transcription system are direct audio-to-tablature capability, as audio is the natural output of the object we're interested in transcribing. Other inputs, like scores or video, are already an extent of transcription that detracts from the impressiveness of a tablature transcriber. Secondly, because PCT examines notes' underlying physical signatures, it can theoretically decode the fretboard position of any note despite how unexpected its location may be. Other methods, like graph searches, rely on optimization of cost functions that are generalizations of tab playability rules, and therefore aren't as well-suited to correctly transcribe aberrant note sequences. Lastly, their method exclusively uses inharmonicity. Other systems that use inharmonicity [1, 3, 9, 19] exploit additional features that aid in string discriminability. We were captivated by the idea of singularly using inharmonicity to achieve impressive transcription scores.

2.2   Inharmonicity Motivation

To understand the motivation for this work, we first need to address the motivation behind inharmonicity as a feature itself. Inharmonicity is key in many string discrimination systems [1, 3, 4, 9, 19]. When an ideal string fixed at both ends is displaced, the restoring force that causes it to oscillate is its tension. However, for a real string with stiffness, its elasticity contributes to this restoring force [11] which causes the frequencies of the modes of vibration to no longer be integer multiples of the fundamental. Instead, they're skewed upward according to:

$$f_k = k f_0 \sqrt{1 + \beta k^2} \tag{2.1}$$

where f_k is the kth harmonic of fundamental f_0 and β is the string's inharmonicity, defined by

$$\beta = \frac{\pi^3 Q d^4}{64 T l^2}. \tag{2.2}$$

7

In words, the inharmonicity β of a vibrating string is a dimensionless quantity that depends on the string's Young's modulus Q, diameter d, tension T, and vibrating length l, and which scales the degree of deviation of the string's kth partial according to equation (2.1). See Figure 2.1. Note that for an ideally harmonic string, $\beta = 0$ and (2.1) reduces to $f_k = kf_0$, aligning with intuition about harmonics' ideal locations in frequency as simply integer multiples of the fundamental.
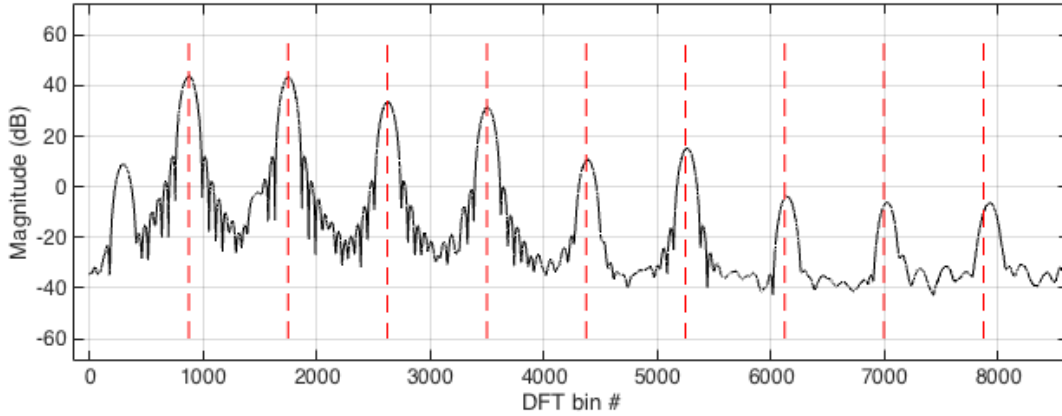
Figure 2.1: Power spectrum of note D2 plucked on an electric guitar at string 5 and fret 5. Red dashed lines are drawn at integer multiples of the fundamental. Note the rightward skew of the measured partial peaks.

Inharmonicity's suitability for the string discrimination task can be understood through equation (2.2). The six strings of any guitar necessarily exhibit different combinations of Q, d, and T due to varying thicknesses, material, and tuning. Each string therefore has an "inharmonic signature" which presumably distinguishes it from the others. Capturing the pattern of a string's inharmonicity variation along its frets should thus reveal sufficiently identifying information.

2.3   Inharmonicity Estimation

Empirical estimation of inharmonicity is itself a topic of research which merits a brief review. Using pitch extraction techniques to perform estimation was first attempted by Galembo in 1979 and 1986 [13, 14]. Several additional methods have been proposed since then. In 1994, Galembo [15] hypothesized a connection between partials-based fundamental pitch estimates and the degree of

inharmonicity in a spectrum. They specifically investigated cepstral analysis and a variant of the harmonic product spectrum method applied to both synthetic tones and recorded piano notes. They found that, because these two techniques leverage periodicity in the frequency domain to produce their pitch estimates, inharmonicity influenced the quality of their estimates in a deterministic manner: more inharmonicity produced wider and less focused pitch estimate lobes, and from quantifying this relation an inharmonicity estimate could be obtained.

The same authors in [16] introduced another method in 1999, dubbed the inharmonic comb filter (ICF) method. To estimate a note's inharmonicity, they first applied sets of comb filters to the note's spectrum. The locations of the comb filters' notches were parametrized with a range of inharmonicity values that sufficiently sampled the note's expected inharmonicity range. The ICF whose inharmonicity best matched that of the note would produce the output with lowest spectral energy, since its notches would most align with the inharmonic note partials. The inharmonicity parameter associated with this output-minimizing ICF was thus selected as the estimate.

Rauhala [26] introduced an efficient procedure in 2007 that iteratively catalogued estimates of the partial frequencies' deviations and returned increasingly better estimates of the inharmonicity. The algorithm, known as the partial frequency deviations (PDF) method, is initialized with a reasonable first estimate of inharmonicity, and then the note's spectrum is searched for partial peaks. Differences are measured between the locations of these discovered peaks and those of the expected peaks, and the aggregate deviation trend is used to refine the inharmonicity estimate – a majority of positive differences implies the inharmonicity should be reduced, and a majority of negative differences implies the opposite.

In 2009, Hodgkinson [18] proposed a yet more efficient and accurate algorithm, named median-adjustive trajectories (MAT). Their routine exploits the fact that inharmonicity can be estimated with any two partials and their corresponding indices in the harmonic series. Pairs of low-index partials are first considered, since they're minimally affected by inharmonic skew and therefore have high location reliability. With these initial inharmonicity estimates, additional partials are collected and used to refine the previous inharmonicity estimates, with which more

9

partials are located, so on and so forth. MAT was the inharmonicity estimation routine employed in this work, and is discussed at length in Chapter 3.

Three years later, Barbancho [4] introduced another PFD-based approach as part of a guitar tablature transcription system. They started by cataloguing the first 10 partials' deviations, then fitting a polynomial to the PFD curve. They derived the relation between the polynomial's coefficients and the inharmonicity, and returned an initial inharmonicity estimate according to this derivation. Then using this inharmonicity to aid in their search for higher-index partials, they catalogued a larger number of partials' deviations than that of the first iteration, and performed the same polynomial fit and inharmonicity derivation routine. This would refine the inharmonicity estimate, allowing for increasing numbers of partials for which to search.

Also in 2012, Abesser [1] used parametric spectral modeling, again in the context of a guitar tablature transcription system. To estimate the inharmonicity of a plucked guitar note, they modeled each audio frame with the autoregressive (AR) filter that best approximated its spectral content. They obtained the poles' locations and the ideal harmonics' locations, and like Barnbancho [4] fit the PFD curve with a polynomial from whose coefficients an inharmonicity estimate could be derived.

Chapter 3:   Characterizing Strings as Inharmonicity Trajectories

In this chapter, we present our automatic tablature transcription system in detail. We begin with the motivation for our approach, then present its various building blocks: the specific inharmonicity estimation algorithm we used, transformation of inharmonicity to log-inharmonicity, learning log-inharmonicity trajectories, compensation for arbitrary tuning, string classification using the learned trajectories, and the final tablature conversion and refinement process.

3.1   Motivation

Barbancho [4] derived that inharmonicity varies in a deterministic manner with respect to fret positions along a given string. Specifically, they showed that the inharmonicity $\beta(s, n)$ of a note produced by string s at fret number n could be expressed in terms of the inharmonicity of the open-string note $\beta(s, 0)$ according to:

$$\beta(s, n) = \beta(s, 0)2^{\frac{n}{6}} \tag{3.1}$$

In other words, the variation in inharmonicity along a given string can be expressed simply in terms of its open-note inharmonicity, effectively defining a restricted trajectory along which the inharmonicity must vary. Figure 3.1 illustrates these trajectories.

Recognizing this, they were able to successfully transcribe string and fret number of an un-
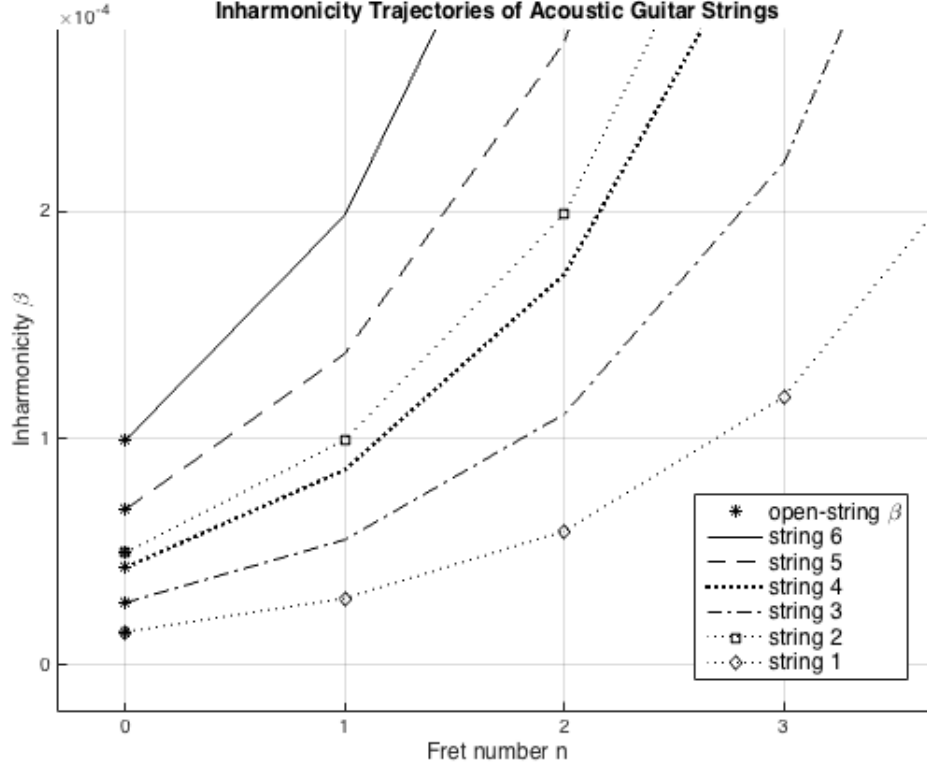
Figure 3.1: Example inharmonicity trajectories of acoustic guitar strings. Only frets 0 (open-string) through 3 are shown. Observe the restriction to the exponential trajectory defined by equation (3.1).

known guitar note. Their system's essential mechanism was a similarity-determination routine $A(X, \hat{\beta}_s)$ that quantified the resemblance between an unknown note's spectrum X and the expected inharmonicities $\hat{\beta}_s$ of candidate strings s from which the note could have originated. First, to estimate an open-note's inharmonicity they followed the procedure described in section 2.3 in which they estimated its fundamental, then catalogued its partials' deviations from their ideal harmonic locations, and fit to them a polynomial from whose coefficients they derived the inharmonicity estimate. Next, they inferred the corresponding fret on which the note could be played for each candidate string, and used these string-fret candidate numbers in equation (3.1) to obtain expected inharmonicity values of the unknown note for each candidate string. In other words, they followed each candidate string's inharmonicity trajectory to arrive at an expected

12

inharmonicity value at the appropriate fret. Lastly, they applied their similarity determination algorithm $A(X, \hat{\beta}_s)$, which went as follows. For each candidate fretboard position, they obtained the expected skewed frequency locations at which to find the inharmonic peaks, and searched the note's spectrum for them using search windows centered at those expected locations. They performed this search for a set number of peaks, all the while cataloguing their deviation from the search window centers. Their motivation was that the expected inharmonicity of the true fretboard position would produce a spectrum with inharmonic peaks that were measurably more coincident with the inharmonic peaks of the unknown note than were any of the imposter strings. They found that, indeed, the inharmonicity of the fretboard position on which the note was truly plucked yielded the largest number of coincident located peaks inside the spectral search window bounds, hence our name "partials coincidence tally" for their algorithm. Their work demonstrated the sufficiency of inharmonicity as a discriminating feature.

This work capitalizes further on their trajectory realization. Rather than zooming into fretboard positions' spectra to evaluate their candidacy, we propose characterizing each candidate string by its inharmonicity trajectory and classifying a note as simply belonging to the string whose inharmonicity trajectory evaluated at the note's fundamental best approximates the note's inharmonicity. In other words, our similarity routine is rather some error calculation $A'(\beta, \hat{\beta}_s)$, discussed in detail in the next sections, between the note's inharmonicity β and each candidate fretboard positions' predicted inharmonicity $\hat{\beta}_s$ for the note. This is an algorithmically simple approach, requiring merely inharmonicity estimation of the unknown note, inharmonicity trends of training guitars' strings, and an error measure. Additionally, this method is governed by the intuitive spatial constraint of inharmonicity trajectory: notes played on a given string can exhibit only a set of inharmonicities, so classifying based on a spatial metric like minimal residual from a regression makes physical sense.

3.2 Inharmonicity Estimation

The first step in our system is estimation of the inharmonicity of an unseen note. The algorithm

we employed was the median adjustive trajectories (MAT) method [18] proposed by Hodgkinson

in 2009, and we present it here. At the core of this method is a recasting of the definition of

inharmonicity in terms of any two partials and their indices. We can rearrange equation (2.1) to

express the fundamental f_0 in terms of the frequency f_m of the mth partial and its index m as

$$f_0 = \frac{f_m}{m\sqrt{1 + \beta m^2}}. \tag{3.2}$$

Next, we can substitute (3.2) back into f_0 in equation (2.1) to obtain

$$f_k = \frac{k f_m}{m\sqrt{1 + \beta m^2}}\sqrt{1 + \beta k^2}. \tag{3.3}$$

Solving for β yields

$$\beta = \frac{(f_k \frac{m}{k})^2 - f_m^2}{k^2 f_m^2 - m^2 (f_k \frac{m}{k})^2}, \tag{3.4}$$

which defines the inharmonicity of a note in terms of an arbitrary two partials k and m and their

corresponding frequencies f_k and f_m.

The MAT algorithm first pre-processes a note with the Complex Spectral Phase Evolution

(CSPE) method [28] to refine the subsequent frequency estimates. The inharmonicity routine

then begins by locating partials 1 and 2 in this refined spectrum based on a user-input estimate

of the fundamental. Searching for these partials is straightforward, since the skew effect of

inharmonicity is low for small indices. From these locations, (3.4) is used for every combination

of the located partials (which for this first iteration, is only one) to obtain an array of β estimates.

The median of the β array is returned as the composite β estimate of this iteration, and this

median together with the partials' locations are used in (3.2) to obtain an array of revised f_0

estimates that contains as many elements as partials under consideration. Finally, the median of

this f_0 array and the median of the β array are used in equation (2.1) to predict the frequency of the next partial for which to search, which is the third in this case. At succeeding iterations, the lengths of these f_0 and β arrays grow longer (since both the number of partials, and therefore the number of combinations of partials pairs, increases), and the estimates which they contain grow increasingly more accurate. The algorithm is summarized in the flowchart in Figure 3.2, taken from [18]. MAT determines the ideal number of partials for which to search by finding the highest partial index that lies above the average level of the magnitude spectrum, and terminates once this partial index is reached. The result is an efficient and accurate inharmonicity estimation routine.

We apply the MAT estimation procedure to five consecutive 100ms windows of audio, starting from detection of a note onset. This yields five inharmonicity estimates which we aggregate into one effective result using the frame aggregation process described in Section 3.6.

3.3 Log-inharmonicity

Next in our pipeline, we apply a log transformation to the estimated inharmonicity to make its variation with respect to MIDI pitch linear. This has the advantage of using simpler regression models for classification at later stages. If we reconsider equation (3.1) in terms of MIDI pitch number m instead of fret number n, we obtain

$$\beta(s,m) = \beta(s,m_{os})2^{\frac{m-m_{os}}{6}},\tag{3.5}$$

where m_{os} is the open-string pitch of string s. The exponential trajectory is maintained since incrementing both fret number and MIDI pitch number accomplishes an equivalent increase in fundamental frequency. Figure 3.3 illustrates the trajectories in terms of MIDI pitches.

15
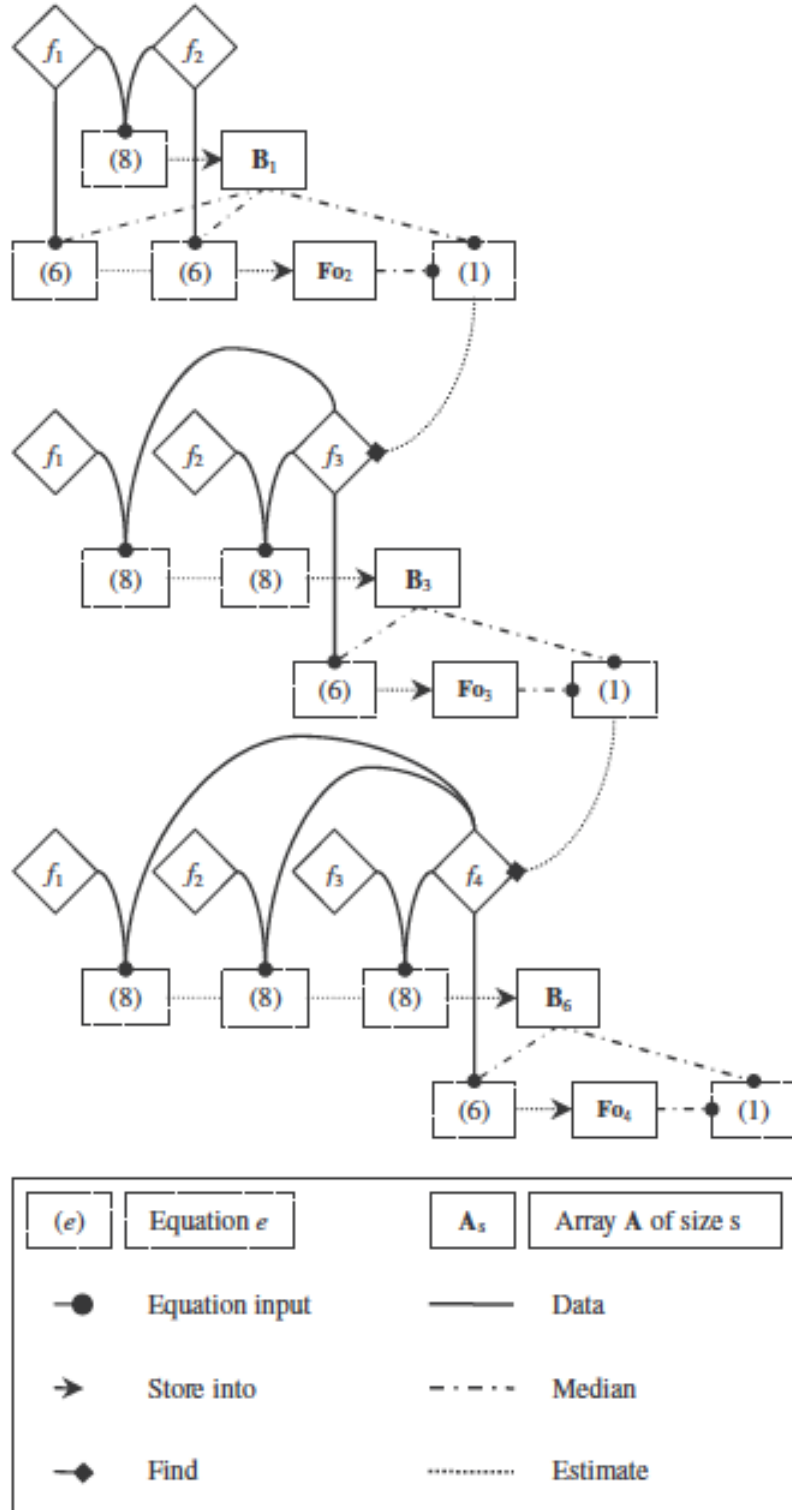
Figure 3.2: Median-adjustive-trajectory (MAT) method flowchart.
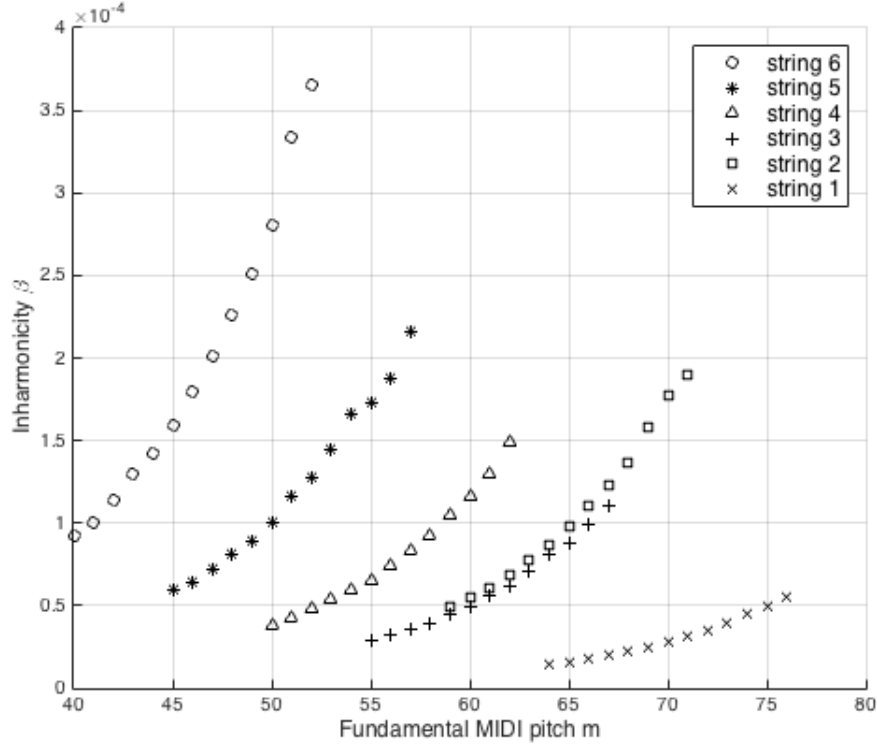
Figure 3.3: Inharmonicity estimates for frets 0-12 on strings 1-6 of RWC acoustic guitar 111AG. Note the exponential trajectory as in Fig. 3.1 and the increased segregation since notes and their inharmonicities are now plotted against their fundamental instead of their fret number.

Now if we consider the log-inharmonicity $\beta_l(m)$, we see that

$$\beta_l(m) = \log_2 \beta(s, m) = \log_2[\beta(m_{os})2^{\frac{m - m_{os}}{6}}] \tag{3.6}$$

$$\beta_l(m) = \log_2 \beta(m_{os}) + \frac{m - m_{os}}{6} \tag{3.7}$$

$$\beta_l(m) = (\log_2 \beta(m_{os}) - \frac{m_{os}}{6}) + (\frac{1}{6})m, \tag{3.8}$$

where string s has been dropped from the notation since we're looking only at variation along a fixed string. Substituting w_0 for $\log_2 \beta(m_{os}) - \frac{m_{os}}{6}$ and letting $w_1 = \frac{1}{6}$, we obtain the familiar linear form

$$\beta_l(m) = w_0 + w_1 m, \tag{3.9}$$

17

showing us that the log-inharmonicity trajectory of a given string varies linearly with respect to MIDI pitch. See Figure 3.4. This relationship also holds for fret number, but fundamental pitch is a more likely and generic feature to have been estimated, and consequently is more relevant to consider here.
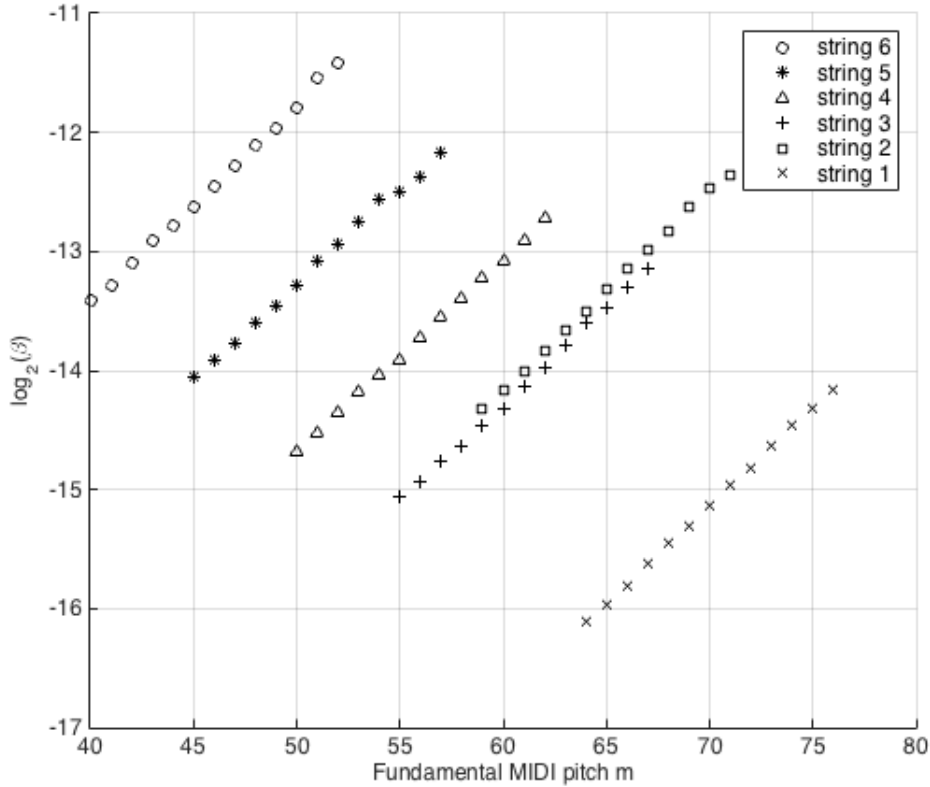
Figure 3.4: Log-inharmonicity estimates for same frets, strings, and guitar as Fig. 3.3

3.4   Learning Inharmonicity Trajectories

We next need to train our system to understand strings' typical log-inharmonicity trajectories. There are two methods we use to do this. The first, which we call *theoretical* trajectory determination, is the method Barbancho employed, so we use this for comparison of our system to theirs. For each guitar, we collect the six notes labeled with fret 0 (i.e. played open-string) from

its recordings, and estimate their inharmonicities. To obtain a more robust estimate, we take the string-wise median of these estimates for our final six open-string inharmonicities. Our line trajectories are then simply those defined by equation (3.9).

The other method we use is linear regressions for obtaining *empirical* log-inharmonicity trajectories. This is one of the novel features of this work. Regressing log-inharmonicities of notes against their pitches could capture a more fitting trajectory than the strictly theoretical trajectories obtained from equation (3.9). In this method, we collect all notes with common string labels in our training data, and we perform linear regression of their log-inharmonicities β against their fundamental pitches m (in MIDI note number format). We do this for each string label. Let $s \in \{1, 2, 3, 4, 5, 6\}$ be the string label to which each training note is assigned, and N_s be the number of notes we have belonging to string label s. If we let $\mathbf{x}_s^{(i)} = [1, m_s^{(i)}]^T$ represent the ith note with fundamental pitch $m^{(i)}$ belonging to string s , and $\beta_s^{(i)}$ be the log-inharmonicity of the ith note belonging to string s, we can solve

$$\mathbf{w}_s = \arg\min_{\mathbf{w}} \sum_{i=1}^{N_s} (\beta_s^{(i)} - \mathbf{w}^T \mathbf{x}_s^{(i)})^2 \tag{3.10}$$

which yields the weight vector $\mathbf{w}_s$ that minimizes the sum of squared error between the measured and predicted inharmonicities of the notes belonging to string s. Figure 3.5 illustrates the differences between the trajectory and regression approaches.

Mistakes in the inharmonicity estimation routine would sometimes produce outliers that deviated substantially from the linear trajectory model, and which disproportionately skewed the regressions. For this reason, we rather performed a bisquare weighted linear regression, instead of the standard unweighted regression in (3.10), to empirically approximate the log-inharmonicities. Bisquare weighting is a robust linear regression procedure that minimizes a weighted sum of squares that de-emphasizes outliers. The iterative procedure works as follows. First, a regular
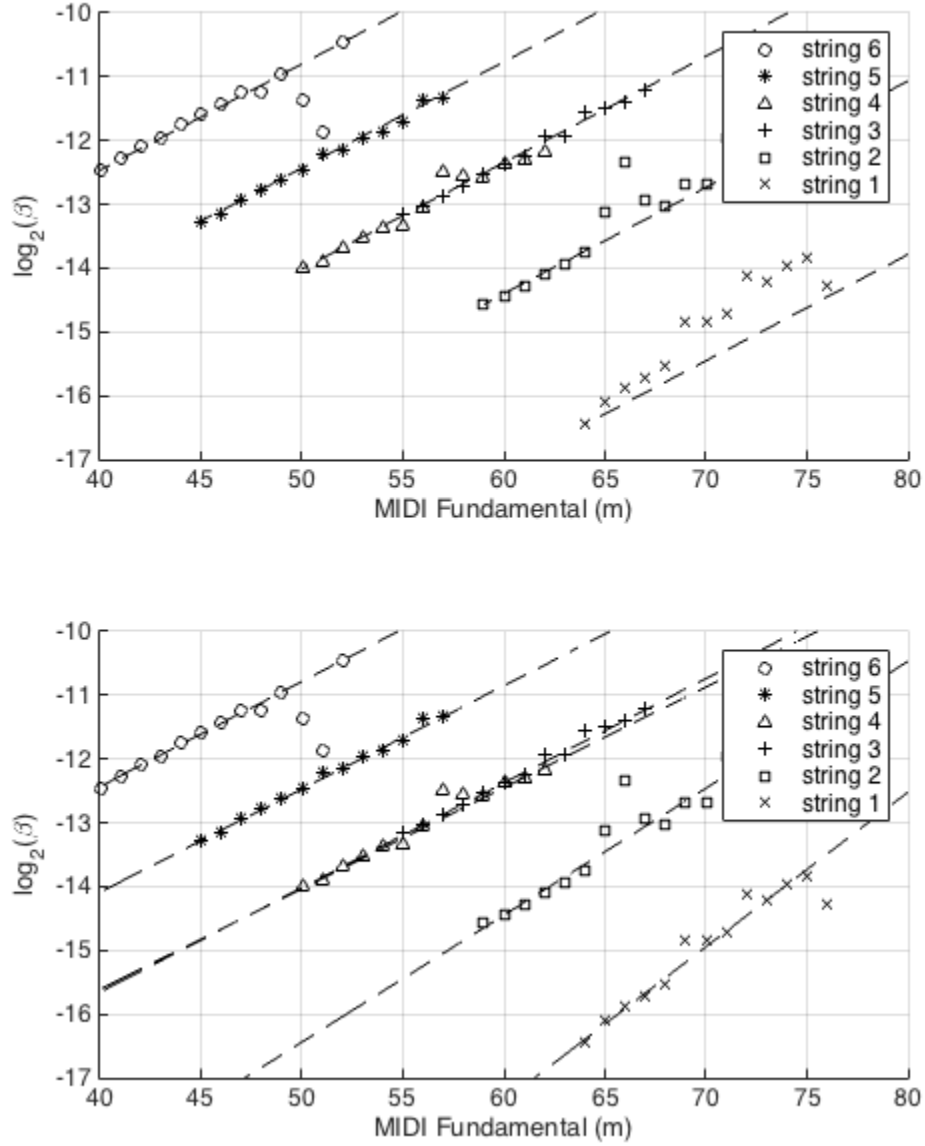
Figure 3.5: Comparison of string inharmonicity characterization approaches for our recorded Fender Telecaster. Top: *theoretical* trajectories. Note how each line begins precisely at the first marker of each string (i.e. the open-string note) and continues along a trajectory that ignores empirical inharmonicity estimates. String 1 most clearly shows this. Additionally, the trajectories for strings 3 and 4 are effectively co-linear and therefore indiscriminable. Bottom: regressions, or *empirical* trajectories. Each regression was obtained by minimizing the bisquare weighted error of the string's residuals, so the lines more closely follow the empirical inharmonicity scatterplot estimates.

least squares fit is obtained. Then, the standardized adjusted residuals are computed according to

$$u = \frac{r_i}{Ks\sqrt{1 - h_i}}, \tag{3.11}$$

where r_i are the plain residuals, h_i are the leverages, $K = 4.685$ is a tuning constant, and s is the robust variance, which is equal to $M/0.6745$ where M is the median absolute deviation of the residuals. Next, the robust weights are computed according to

$$w_i = \begin{cases} (1 - u_i^2)^2, & |u_i| < 1 \\ 0, & |u_i| \geq 1 \end{cases} \tag{3.12}$$

The process is repeated, iteratively improving the weights until the fit converges [29]. The Matlab 'robustfit' routine was used to implement this.

3.5 Tuning Compensation

After transforming our data into log-inharmonicity and MIDI pitch space, we apply a tuning compensation feature step if the test guitar is known to be in alternate-tuning. In this section, we derive the straightforward adjustment and explain how it is incorporated into our system.

Though standard tuning is the most common pitch configuration of six-string guitars, there exist numerous other tunings in which performers often play. Aside from altering the musicality of the instrument, alternate tunings complicate the transcription process by introducing uncertainty about the open-string pitches. They distort inharmonicity-based methods, since the tensions T of alternately-tuned guitar strings differ from those in standard-tuned strings, thereby affecting estimation of the inharmonicity.

Current string classification systems don't address this degree of freedom. PCT [4] is technically tuning-invariant, but would require access to recordings of the alternately-tuned test guitar's open strings and so isn't considered a classifier here. Other supervised learning systems [1, 9, 19]

are trained and evaluated only on standard-tuning guitars.

A possible approach to augment inharmonicity-based string classifiers with tuning-invariant performance is to simply introduce a scaling factor on the expected inharmonicity, or equivalently an additive factor on the expected log-inharonicity. The fundamental frequency f_0 of an ideal vibrating string is related to its tension T according to

$$f_0 = \frac{1}{2L}\sqrt{\frac{T}{\mu}} \tag{3.13}$$

where L is string length and μ is its density. From this, we can see that

$$T \propto f_0^2. \tag{3.14}$$

Recognizing that for a change in fundamental pitch of Δm semitones, the equivalent change in frequency is $2^{\frac{\Delta m}{12}}$, we see that the proportional change in tension (with all other factors constant) is

$$T \propto 2^{\frac{\Delta m}{6}} f_0^2. \tag{3.15}$$

The resulting log-inharmonicity of this new open-string note with pitch $m'_{os} = m_{os} + \Delta m$, with original open-string pitch m_{os}, is therefore

$$\log_2 \beta(m'_{os}) = \log_2 \left[\frac{\pi^3 Q d^4}{64 T l^2} (2^{-\frac{\Delta m}{6}}) \right]. \tag{3.16}$$

$$\log_2 \beta(m'_{os}) = \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{\Delta m}{6}. \tag{3.17}$$

Substituting (3.17) into w_0 from equation (3.9) to obtain the new intercept term w'_0, we get

$$w'_0 = \log_2 \beta(m'_{os}) - \frac{m'_{os}}{6} \tag{3.18}$$

$$w'_0 = \left(\log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{\Delta m}{6} \right) - \frac{m_{os} + \Delta m}{6} \tag{3.19}$$

$$w_0' = \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{m_{os}}{6} - \frac{\Delta m}{3} \tag{3.20}$$

$$w_0' = w_0 - \frac{\Delta m}{3}. \tag{3.21}$$

The slope coefficient $w_1 = -\frac{m}{6}$ is independent from Δm, so the affected log-inharmonicity trajectory of this alternately-tuned string is therefore

$$\beta'(m) = w_0' + w_1' m \tag{3.22}$$

$$\beta'(m) = w_0 + w_1 m - \frac{\Delta m}{3}. \tag{3.23}$$

In words, equation (3.23) tells us that the effect of a string's alternate tuning is simply addition of a bias term $\frac{-\Delta m}{3}$ to the log-inharmonicity, where Δm is the semitone deviation from standard tuning. We should thus be able to compensate our predictive trajectories to arbitrarily-tuned guitars, provided we're given the alternate tuning in which the unseen notes are being performed. As can be seen in Figure 3.5 when one string on the standard-tuned electric guitar is modified using equation (3.23), the predictive alternate-tuning regression aptly captures the trend of the alternate-tuning inharmonicity. This is a straightforward modification for any inharmonicity-exploiting classifier, and we report our results with this modification in the next chapter.

3.6   String Classification

The next step is to use these learned string trajectories to classify unknown notes. Since these trajectories tell us how to expect different strings' inharmonicities to vary with respect to pitch, we think of string classification in this context as simply determining which trajectory or line best approximates the unseen note's inharmonicity estimate. To this end, we simply assign a note to the trajectory that minimizes the reconstruction error.

Other decision functions were considered before deciding on residual minimization. We
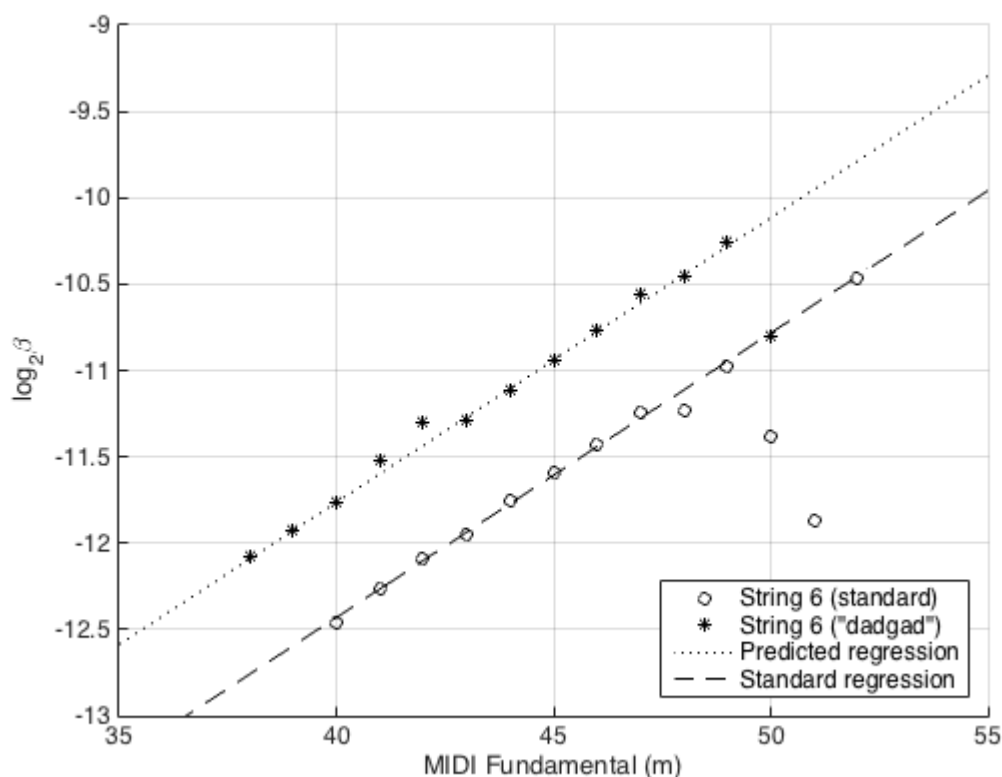
Figure 3.6: Effect of alternate tuning on an electric guitar's log-inharmonicity trajectories. The circles are log-inharmonicity estimates of frets 0-12 on string 6 of RWC electric guitar EG131 in standard tuning, and the dashed line is their bisquare-weighted regression. The asterisks are log-inharmonicity estimates for the same frets, string, and guitar, but tuned two semitones down. The dotted line passing through them is the *predicted* regression obtained by applying equation (3.23) to the dashed line.

considered using a two-dimensional Euclidean distance metric, but ultimately steered away because it wasn't true to the physical phenomenon – expected inharmonicity on a fixed string should be a function of only one variable, namely pitch, so incorporating an additional degree of freedom seems inappropriate. Indeed, one can easily imagine situations where the Euclidean distance between a note's inharmonicity estimate and an erroneous string is smaller than the one-dimensional residual between it and its true string, thereby producing an incorrect string classification.

We also considered a probabilistic decision function. Because, for linear regression, minimizing the residual sum-of-square error of the data is equivalent to maximizing the likelihood of the data if its assumed the data was generated by Gaussian distributions with means centered at the regression line, we also attempted a probabilistic, or "soft" residual classifier. Unseen notes were assigned to the line from which it was most probably generated, given the lines' means and variances. The term "soft" arises from the fact that points are no longer necessarily assigned to the closest line – the decision is evaluated strictly on probabilistic terms. We ultimately avoided this method, however, because our system was suffering from precision issues. The variances of the robust weighted regressions would consistently be tens or hundreds of orders of magnitude smaller than the distances to some outlier inharmonicities. The system would then incorrectly classify these outliers because of insufficient resolution. A workaround wasn't implemented because this, too, didn't quite fit the mechanism by which these inharmonicities were being produced in reality. Inharmonicity outliers were strictly due to inharmonicity estimation mistakes, not a physical phenomenon in the guitar, so it didn't seem appropriate to incorporate this into our model.

We thus settled on minimizing the residual error, which is explained here. We take the nth note in our test set, characterized by tuple $(m^{(n)}, \beta^{(n)})$ where $m^{(n)}$ is its MIDI pitch number and $\beta^{(n)}$ is its inharmonicity. We'd like to transform this into another tuple $(s^{(n)}, f^{(n)})$ corresponding to the appropriate string and fret to which this unknown note should be assigned. We straightforwardly assign to $s^{(n)}$ the index j of the linear coefficients $\mathbf{w}_j$ whose predicted inharmonicity $\hat{\beta}_j$

approximates the measured inharmonicity $\beta^{(n)}$ with least error. More formally, we solve

$$s^{(n)} = \arg\min_{j} (\beta^{(n)} - \hat{\beta}_j)^2 \tag{3.24}$$

where $\hat{\beta}_j$ is the inharmonicity prediction of string j, given by

$$\hat{\beta}_j = \mathbf{w}_j^T \mathbf{x}^{(n)}. \tag{3.25}$$

Here, $\mathbf{x}^{(n)} = [1, m^{(n)}]^T$ to account for a y-intercept term. Figure 3.6 graphically illustrates this process.
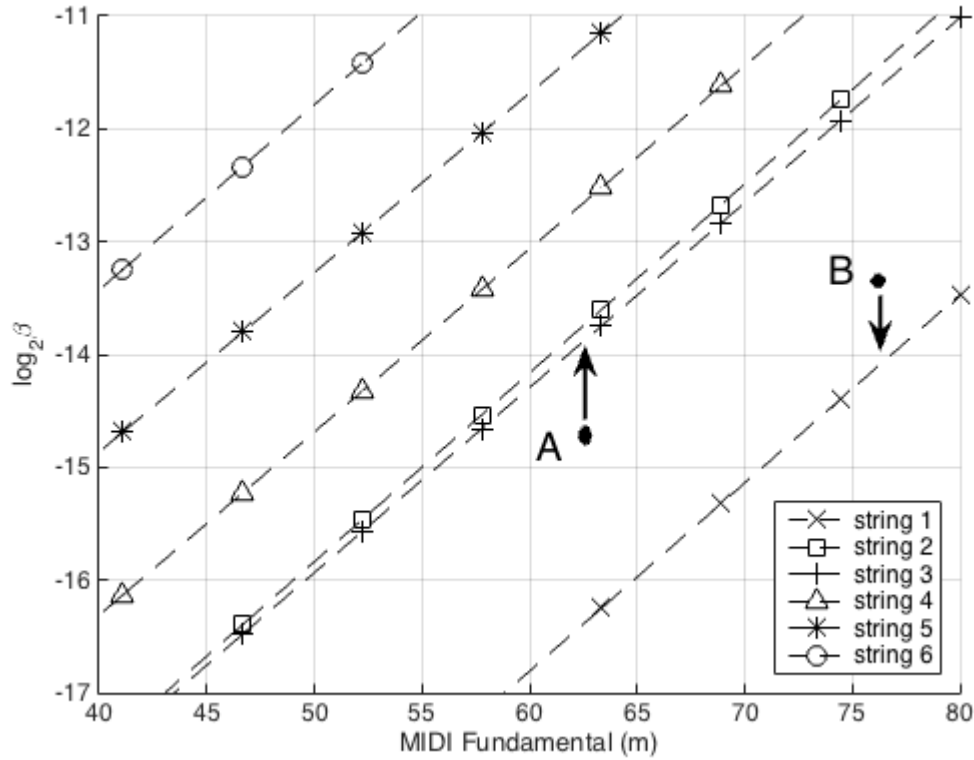
Figure 3.7: Classifying two new notes A and B using learned log-inharmonicity characterizations. Dashed lines are actual learned regressions for RWC acoustic guitar AG091. Note A with fundamental pitch 62 is assigned to string 3, since its line characterization is closest to A's inharmonicity estimate. The residual of note B, with fundamental pitch 76, is smallest with respect to string 1, so it's assigned to that string under the least-square-error decision of equation (3.24).

Occasionally, outlier inharmonicity estimates would throw off this classification scheme. To combat this, we employed frame aggregation, a refinement technique used by [1]. In frame aggregation, inharmonicities of multiple audio windows are estimated and classified individually, then the multiple results are aggregated into a compound final decision. In essence, it's a form of averaging that relies on the theoretically low probability that consecutive inharmonicity estimates will all be outliers; the more probable stable estimates will contribute to a more reasonable aggregate string decision. For each frame r in an F-frame aggregation of a note, we obtain the string-wise distances $\mathbf{d_r} = [d_1, d_2, d_3, d_4, d_5, d_6]^T$ between its log-inharmonicity estimate and all possible strings s. Then, we simply sum the distances $\mathbf{d_r}$ over all frames and select the string whose aggregate distance is minimal:

$$\hat{s} = \underset{s \in \{1,2,3,4,5,6\}}{\arg\min} \sum_{r=1}^{F} \mathbf{d_r}. \qquad (3.26)$$

The result is a tamer inharmonicity estimate and a slight boost in classification accuracy. For our system, we used 5 abutting frames, each 100ms long, which start at the detected onset of a note. Onsets were obtained through automated detection, the details of which are beyond the scope of this work but are cited for reference [5, 10].

3.7   Tablature Conversion and Refinement

The final step is transformation of the string classifier output into tablature notation. Because the classifier outputs are string numbers, we still need to infer notes' fret positions. This is a trivial task provided we know the tuning $\mathbf{t} = [m_1, m_2, m_3, m_4, m_5, m_6]^T$ of the unknown guitar, where m_s is the MIDI pitch number of the open note on string s. Because we already have an estimate of the nth unknown note's MIDI pitch $m^{(n)}$, simply taking the difference between $m^{(n)}$ and its assigned string's open pitch m_s yields the fret on which the note was played; each fret on a guitar increases the string's pitch by one half-step, or equivalently the MIDI pitch number by

one integer.

We further refine the fretboard estimates with "plausbility filtering" [1] in which we reject and reassign those string decisions that are implausible given the constraints and assumptions of the data. If our systems assigns to a note a string on which the corresponding fretted position is negative, a mistake has clearly been made since frets can't possibly be lower than 0, or the open-string. Similarly, if our system assigns to a note a string on which the corresponding fret is greater than 12, we know this is an error as well because our guitar recordings feature only frets 0-12. When either of these situations arise, we reject the string classification decision and select the next closest string. If the succeeding string is also implausible, the process is iterated until all six strings have been considered.

3.8   System Overview

Below in Figure 3.8, we summarize the operation of our system. Inharmonicity estimates are obtained for our training guitars' notes, then we represent these estimates in log-inharmonicity space with respect to MIDI pitch to obtain linear trends. We then check if the evaluation guitars' tunings are non-standard, and if so, we appropriately compensate the weight vectors that define the linear trajectories. Next, we perform classification on evaluation recordings by assigning to unknown notes the strings whose log-inharmonicity trajectories minimize the residual errors of the notes' log-inharmonicities. Finally, tablature conversion and refinement is performed with plausibility and constraint filtering.
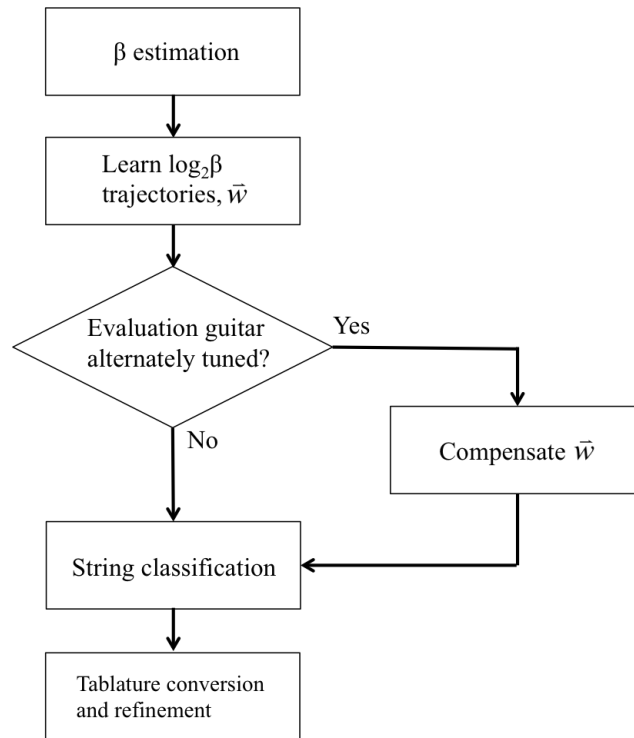
Figure 3.8: System overview of tablature transcription with log-inharmonicity trajectories using residual minimization classification.

Chapter 4:   Experiments and Results

4.1   Benchmark RWC Evaluation

We used subsets of the Real World Corpus Music Instrument Database (RWC-MDB-I) [17] and personal recordings of an electric and acoustic guitar to conduct our experiments.

Our RWC subset comprised nine guitar recordings (three classical, three acoustic, and three electric), each of which were performed twelve times exhibiting various permutations of particular musicality parameters (playing style, dynamic level, pickup selection). For our work, we used only six of the twelve recordings for each of the electric guitars, as the discarded ones featured musicality parameters like vibrato and palm-muting that weren't featured in the classical and acoustic sets. The performances themselves were simply clean, isolated, monophonic enumeration of every fret (from open-string to 12th fret) on every string (from string 6 to string 1), constituting 78 total note plucks per audio file. The resolution is 16 bits per sample, at 44.1kHz. Labels of strings and frets are thus obtained by their location of occurrence in the recording. The classical guitars recorded were a Stafford, a Sakurai Kohno Professional-J, and a Yuichi Imai YJ-II; the three acoustic guitars captured were an Ovation, a Yamaha APX, and Yairi WY1; the electric guitars featured were a Fender Stratocaster, an Aria PE, and an Ibanez Artcore.

We also recorded a Fender Telecaster (electric guitar) and a Takamine G-series (acoustic guitar) at 16 bits and 44.1kHz in the same vein as the RWC database: string and fret enumeration, totaling 78 notes per recording. We captured various tunings: standard, "DADGAD" (in which the open pitches of strings 6 through 1 are given by the respective letters in name), "WSU"

(whole-step up), and "WSD" (whole-step down). The Fender, which was recorded directly into our audio interface, was performed with a thick plectrum at center pickup orientation and at similar moderate dynamic levels throughout. We recorded the Takamine in an isolated acoustic chamber with a large diaphragm condenser microphone (an MXL 990) placed 12 inches away from fret 12, and we also played it with a thick plectrum and at constant dynamic levels. Analog-to-digital hardware included a Focusrite Saffire Pro 24 and a 2011 Macbook Pro running Logic. No other musicality parameters were captured; we focused on tuning variance here.

We conduct experiments on all nine RWC guitars at three levels of context, which comprise the headings of Table 4.1: guitar-specific, guitar-averaged, and guitar-independent. For guitar-specific trials, we train and test the system on the same guitar using three-fold cross-validation with test folds that are one-third of the guitar's recordings. The folds' results are averaged together for composite transcription errors. For guitar-averaged trials, we train the system on all recordings of the two guitars in the same class as the test guitar, as well as a two-thirds training portion of the test guitar, and then evaluate on the remaining third of the test guitar recordings. Again, three-fold cross-validation is used to iterate through the portions of the test recordings. For guitar-independent trials, the system never sees any of the test guitar's recordings before evaluation. We train the system on two of the guitars in the same class as the test guitar, and test on all recordings of the remaining one to obtain its transcription performance. In all cases, Table 4.1 reports error probabilities, as this was the transcription metric used in [4].

For our first experiment, we evaluated string identification performance of a rudimentary maximum-likelihood classifier as a baseline. Because currently the only standard for solely inharmonicity-based transcription performance is that of PCT, it's essential that we interpret any performance gains with our proposed method in light of an agnostic naïve baseline method. To this end, inharmonicities from the training audio were collected by fretboard position, and their medians were calculated. To classify an unknown note, we simply obtained its candidate fretboard positions and returned that which maximized the probability of having observed the note's measured inharmonicity. Because the inharmonicities often didn't closely follow a normal

distribution, we used as a proxe for probability the inverse squared-distance to the candidate fretboard positions' median inharmonicities. Results from this baseline experiment, conducted for all three evaluation scenarios discussed in the previous paragraph, are displayed under the "MLE" columns of Table 4.1.

We next report string identification results using the PCT method, taken from [4]. These are located under the "PCT" columns of Table 4.1. As discussed in Chapter 2, this method uses estimates of the open-string inharmonicities of the training audio's guitar(s) to predict *expected* inharmonicities of an unknown note's candidate fretboard positions. The degree of similarity between the unknown note's measured inharmonicity and that of the candidate fretboard positions is assessed by tallying their coincident partials. The candidate fretboard position that produces the highest tally is returned as the transcribed tablature. Note that no results for the "Guitar-independent" column exist for this method, as no results were reported in [4] for this scenario.

Finally, we evaluate our trajectory residual method. We performed bisquare weighted regression of the training guitars' log-inharmonicity estimates against their pitches. String classification was performed on the test notes using residual minimization on the regressions, as discussed in Chapter 3. Error probabilities are reported under the "Regr." columns of Table 4.1. We also evaluated our *theoretical* trajectory method, in which regressions aren't computed but residual minimization is still used for classification. However, error probabilities were only slightly better than those of our *empirical* trajectory (i.e. regression) method, so they're omitted from Table 4.1 for conciseness.

We point the reader's attention to the "Overall..." rows in Table 4.1 to highlight the central results of this work. For the columns under the "Guitar-specific" column, PCT excels for every guitar. Clearly, tallying the coincident partials using inharmonicity estimates from the exact guitar on which tests are conducted outperforms the MLE baseline and our regression-residual method. Though surprisingly for the columns under the "Guitar-averaged" and "Guitar-independent" heading, the straight-forward baseline MLE performs best, and is nearly unwavering in accuracy throughout the varying experimental scenarios. Also noteworthy is our regression-residual

method's superior performance to PCT's for acoustics and classicals; it seems that capturing

the training audio's inharmonicity trajectories rather than just its open-string inharmonicities is

a more robust transcription scheme when the test guitars increasingly differ from those in the

reference set.

Transcription Error Probabilities									
	Guitar-specific			Guitar-averaged			Guitar-independent		
Guitar	MLE	PCT	Regr.	MLE	PCT	Regr.	MLE	PCT	Regr.
Classical CG091	0.02	0.04	0.03	0.13	0.09	0.03	0.14	–	0.03
Classical CG092	0.15	0.00	0.06	0.16	0.01	0.06	0.13	–	0.03
Classical CG093	0.08	0.01	0.04	0.14	0.00	0.06	0.14	–	0.06
Overall CG:	**0.08**	**0.02**	**0.04**	**0.14**	**0.03**	**0.05**	**0.14**	–	**0.05**
Acoustic AG111	0.00	0.00	0.03	0.01	0.06	0.03	0.02	–	0.03
Acoustic AG112	0.02	0.00	0.03	0.03	0.22	0.06	0.13	–	0.13
Acoustic AG113	0.03	0.00	0.04	0.06	0.04	0.06	0.10	–	0.07
Overall AG:	**0.02**	**0.00**	**0.03**	**0.03**	**0.11**	**0.05**	**0.08**	–	**0.08**
Electric EG131	0.22	0.00	0.18	0.21	0.20	0.18	0.24	–	0.23
Electric EG132	0.15	0.01	0.10	0.15	0.23	0.17	0.17	–	0.21
Electric EG133	0.06	0.00	0.09	0.20	0.32	0.19	0.22	–	0.27
Overall EG:	**0.14**	**0.00**	**0.12**	**0.19**	**0.25**	**0.18**	**0.21**	–	**0.24**

Table 4.1: Error probability comparisons between PCT, the baseline maximum likelihood esti-
mate classifier MLE, residual minimization with theoretical trajectories, and residual minimiza-
tion with empirical trajectories. PCT results are taken from [4]; we've excluded error due to f_0
estimation in their results.

Also displayed in Tables 4.2, 4.3, and 4.4 are the *string-wise* F1-scores using our regression-

based classifier for the "Guitar-averaged" trial. The F1-score is the harmonic mean of precision

and recall, and is a commonly used classification metric elsewhere, so we use it here and wher-

ever comparison with [4] is absent. These tables' left-hand columns denote the string number

of the rows' F1-scores, and the right-hand columns are the row-wise average of the three guitars

in the class. We present this information to highlight the individual strings' contributions to the

overall transcription accuracy.

Transcription F1-scores			
String No.	CG1	CG2	CG3	Overall CG
1	0.96	0.93	0.96	0.95
2	0.92	0.90	0.91	0.91
3	1.00	0.95	0.96	0.97
4	0.96	0.93	0.91	0.93
5	1.00	0.96	0.93	0.96
6	1.00	0.98	1.00	0.99

Table 4.2: String-wise F1-scores for the RWC classical guitars, using regression-residual minimization for the "Guitar-averaged" scenario.

Transcription F1-scores			
String No.	AG1	AG2	AG3	Overall AG
1	0.96	0.96	0.96	0.96
2	0.94	0.80	0.87	0.87
3	0.91	0.89	0.85	0.88
4	1.00	1.00	0.99	1.00
5	1.00	1.00	0.99	1.00
6	1.00	1.00	1.00	1.00

Table 4.3: String-wise F1-scores for the RWC acoustic guitars, using regression-residual minimization for the "Guitar-averaged" scenario.

Transcription F1-scores			
String No.	EG1	EG2	EG3	Overall EG
1	0.96	0.94	0.96	0.95
2	0.94	0.95	0.72	0.87
3	0.47	0.42	0.42	0.44
4	0.66	0.72	0.79	0.72
5	0.85	0.92	1.00	0.92
6	0.96	0.98	1.00	0.98

Table 4.4: String-wise F1-scores for the RWC electric guitars, using regression-residual minimization for the "Guitar-averaged" scenario.

4.2 Tuning Compensation

For the next experiment, we learned inharmonicity regressions from all recordings of the standard-tuned RWC electric and acoustic guitar recordings, and predicted string classes on our alternate-tuned personal recordings. We evaluated string-wise F1-score performance both with and without the tuning compensation factor in our system discussed in Chapter 3. Tuning compensation results are denoted by the "comp." column, and original results without compensation are denoted by the "orig." column. See Table 4.5. For reference, we also report transcription performance on our Fender Telecaster and Takamine G-series in standard tuning, in the second column from the left.

In standard tuning, strings 6 through 1 are tuned to E2, A2, D3, G3, B3, and E4. In so-called "DADGAD" tuning, the open-string pitches are made to be D2, A2, D3, G3, A3, and D4 as these open pitches concatenated together form its name. Whole-step up and whole-step down are tunings in which each of the standard-tuned strings are respectively incremented or decremented by two semitones. For whole-step down, this results in D2, G2, C3, F3, A3, and D4. For whole-step up, we have F#2, B2, E3, A3, C#4, and F#4.

Alternate-tuning Transcription F1-scores (Electric)							
	Standard	"DADGAD"		"WSU"		"WSD"	
String No.	orig.	orig.	comp.	orig.	comp.	orig.	comp.
1	0.96	0.75	0.96	0.83	0.96	0.80	0.96
2	0.96	0.78	0.92	0.58	0.96	0.83	0.96
3	0.50	0.40	0.43	0.56	0.50	0.29	0.50
4	0.76	0.76	0.74	0.23	0.76	0.65	0.76
5	1.00	1.00	0.87	0.33	1.00	0.85	1.00
6	1.00	0.92	1.00	0.88	1.00	0.85	1.00
Overall:	0.86	0.77	0.82	0.57	0.86	0.71	0.86

Table 4.5: Best string-wise F1-scores for our electric Fender Telecaster at various tunings. Regressions trained on all recordings of all three RWC electric guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The "Overall" row displays average F1-score of each column.

Alternate-tuning Transcription F1-scores (Acoustic)							
	Standard	"DADGAD"		"WSU"		"WSD"	
String No.	orig.	orig.	comp.	orig.	comp.	orig.	comp.
1	0.96	0.82	0.92	0.92	0.96	0.82	0.96
2	0.82	0.70	0.63	0.45	0.82	0.61	0.92
3	0.96	0.84	1.00	0.71	0.96	0.00	0.96
4	0.87	0.79	0.79	0.81	0.87	0.18	0.93
5	1.00	1.00	1.00	0.77	1.00	0.61	1.00
6	1.00	0.92	1.00	0.83	1.00	0.81	1.00
Overall:	0.93	0.84	0.89	0.75	0.93	0.50	0.96

Table 4.6: Best string-wise F1-scores for our acoustic Takamine G-series at various tunings. Regressions trained on all recordings of all three RWC acoustic guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The "Overall" row displays average F1-score of each column.

Chapter 5:   Discussion

In this chapter, we begin with an analysis of the previous chapter's results, followed by consideration of factors that potentially limit this work's significance. We close with a discussion about future work.

5.1   Analysis

The baseline MLE approach outperformed both the existing PCT method and our new regression method. This was surprising; we expected that our regression method, by exploiting the underlying linear structure of the log-inharmonicity trajectories, would more accurately predict unseen notes' inharmonicities. Also intriguing is its robust performance across experimental scenarios; classical and acoustic transcription performance for the baseline MLE is constant throughout the "Guitar-specific", "Guitar-averaged", and "Guitar-independent" contexts, with electric transcription performance degrading only marginally for the "Guitar-independent" case.

Our regression-residual approach, though outdone by the naïve MLE classifier, still exhibits noteworthy behavior. We performed worse than PCT for the case when training and testing were both performed exclusively on one guitar whose open-string log-inharmonicities we had access to (the "Guitar-specific" heading in Table 4.1, comparing columns "PCT" and "Regr."), but we achieve lower error probabilities for the most part when the training set diversifies and multiple guitars' log-inharmoncities are considered (the "Guitar-averaged" heading). It seems this is a strength of our system; PCT shines with reliable specific test-guitar inharmonicity information,

but degrades faster and more variably when the reliability of that information decreases. Its worst case performance degradation (for EG133) is from zero error to 25% error as the guitar-specific inharmonicity context is lost, but the error probability of our trajectory residual classification, on the other hand, worsens only marginally at its worst case (also for EG133), degrading by only 8 percentage points. PCT performance on guitar-averaged inharmonicities is also more erratic, with error probability variances for all three guitar classes about one order of magnitude greater than our residual classification. It appears that the residual classification approach trades peak performance capability for more stable results that are also more robust to different guitars' inharmonicities.

The alternate-tuning experiments validated our tuning compensation feature nicely. We see in Tables 4.5 and 4.6 that the general effect of appropriately scaling the inharmonicity regressions is to restore transcription F1-scores towards best-case standard-tuned performance.

Further insight into the error probabilities in Tables 4.1 is gained by considering the string-wise F1-scores in Tables 4.2, 4.3, and 4.4. We see that across all guitars, strings 2, 3, and 4 consistently exhibit lower F1-scores. Strings 1, 5, and 6, contrastingly, exhibit exceptional performance, so we know that any poor results in the "Regr." column in Table 4.1 stem from poor accuracy on these problematic "middle" strings that are physically surrounded on the instrument by higher and lower strings. We thus see a flaw in this residual-based inharmonicity similarity approach: because of its inherently spatial classification decision (residual minimization), this system exhibits poor performance on strings whose trajectories share overlapping regions in inharmonicity-pitch space.

It's unclear why performance for electric guitars is invariably worse than for classicals and acoustics. To investigate, we visually compared our string-wise inharmonicity regressions of all guitars in each guitar class. As can be seen in Figures 5.1, 5.2, and 5.3, the classical and acoustic guitars' inharmonicity regressions remain fairly stable from guitar to guitar, whereas those of the electric guitars clearly exhibit more variability, making residual-based classification more difficult. Electric guitar regressions belonging to strings 3 and 4 overlap considerably,

	Classification					
Truth	1	2	3	4	5	6
1	**0.92**	0.08	0	0	0	0
2	0	**0.92**	0.08	0	0	0
3	0	0.08	**0.38**	0.54	0	0
4	0	0	0.08	**0.62**	0.30	0
5	0	0	0	0	**0.77**	0.23
6	0	0	0	0	0	**1.00**

Table 5.1: Confusion matrix for one classification trial with EG131. Strings 3 and 4, as they are here, were frequently misclassified because of their regressions' poor discriminability.

which is a behavior not absent in the classicals (strings 1 and 4, 2 and 5, 3 and 6) and acoustics (strings 2 and 3) with some strings but which manifests in less error due to their regressions' more consistent placements than those of their electric counterparts. Indeed, the electric guitars' confusion matrices support this, and one which highlights this error trend is shown in Table 5.1.

Further analysis revealed that inconsistent inharmonicity estimates were responsible for the electric guitars' regression variability. We examined the quality of fit of string regressions obtained from all recordings of each of the 3 guitars from each of the 3 types, and found that the classicals' and acoustics' range of determination coefficients was substantially smaller than the electrics'. Highest and lowest quality regressions for CG091, AG111, and EG131 are shown in Figure 5.4 as an example of this; CG091's and AG111's best and worst strings' r^2 coefficients differ by only about one-hundredth, but the r^2 difference for EG131 is an enormous 0.87.

Speculating why this might be the case, we wondered if the transduction of the EG pickups somehow played a role in impeding reliable inharmonicity estimation – either through their transfer functions, possible noise introduction, or some non-linear characteristic. If this were the case, it would follow that higher-intensity performances would elicit this impeding phenomenon to a greater extent and produce measurably more variable inharmonicity estimates than would lower-intensity performances. Since one of the RWC musicality parameters that varies for each featured guitar is dynamic level of the performance, we were able to separate higher- and lower-intensity performances of the same guitar and analyze this interaction between degree

Figure 5.1: Superimposed string-wise regressions of all three RWC classical guitars. Top: strings 1, 2, and 3. Bottom: strings 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the consistent intercept and slope despite varying guitar (with the only exception being string 6).
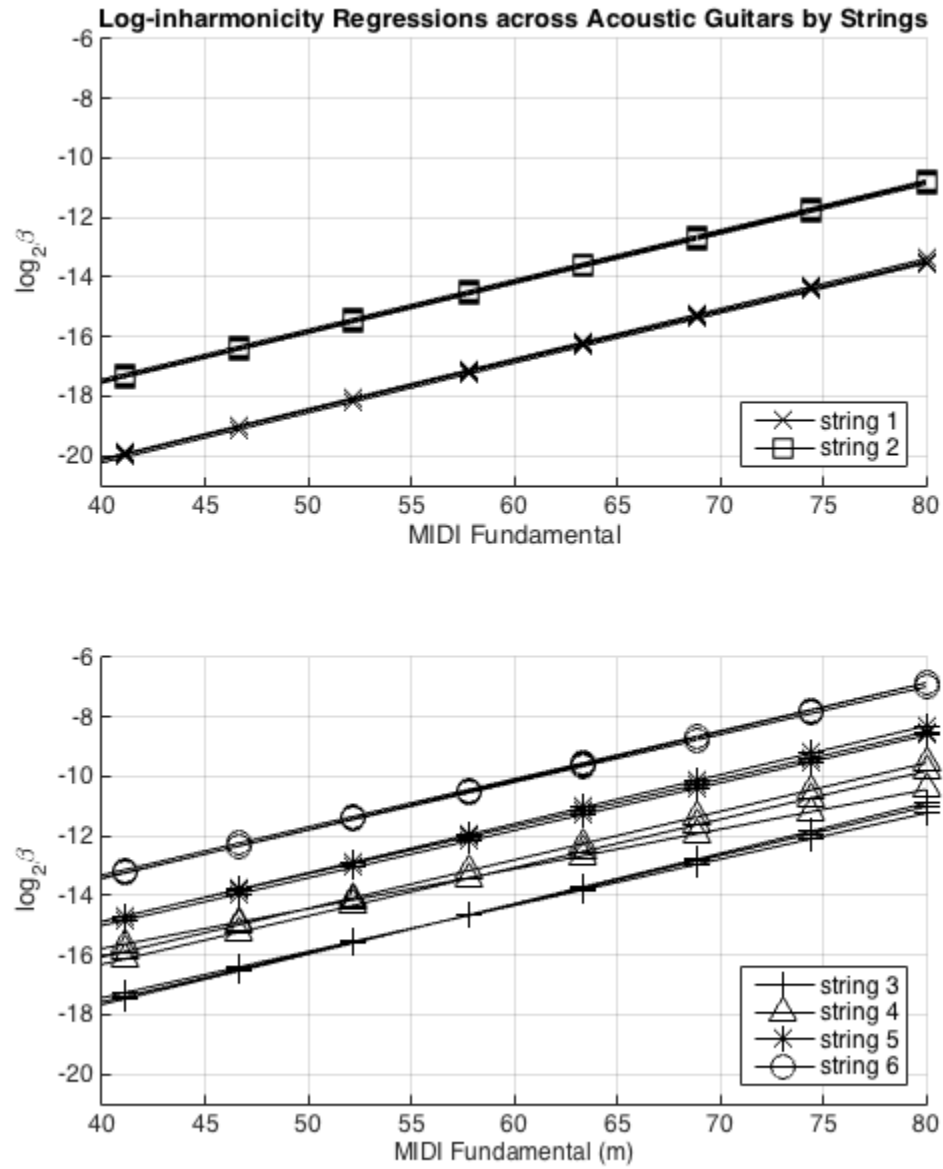
Figure 5.2: Superimposed string-wise regressions of all three RWC acoustic guitars. Top: strings 1 and 2. Bottom: strings 3, 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the consistent intercept and slope despite varying guitar.
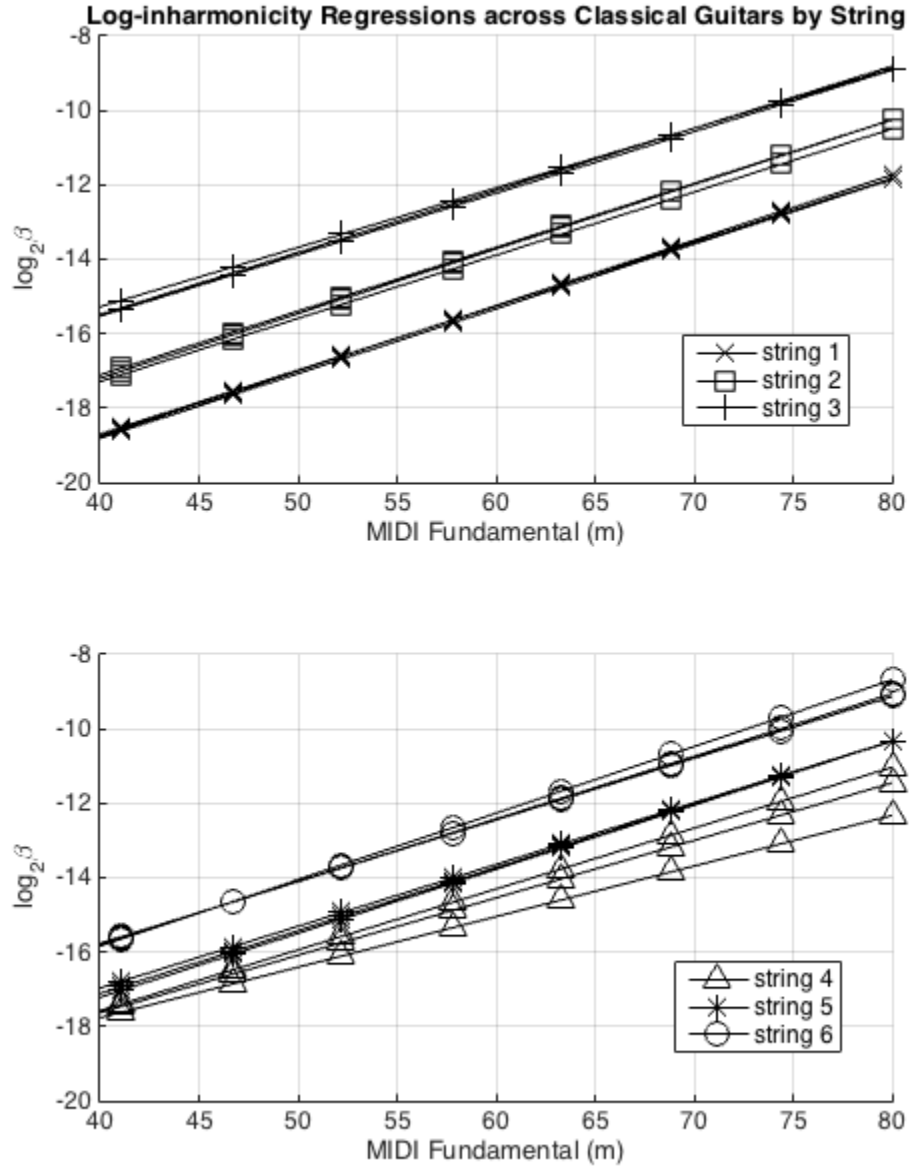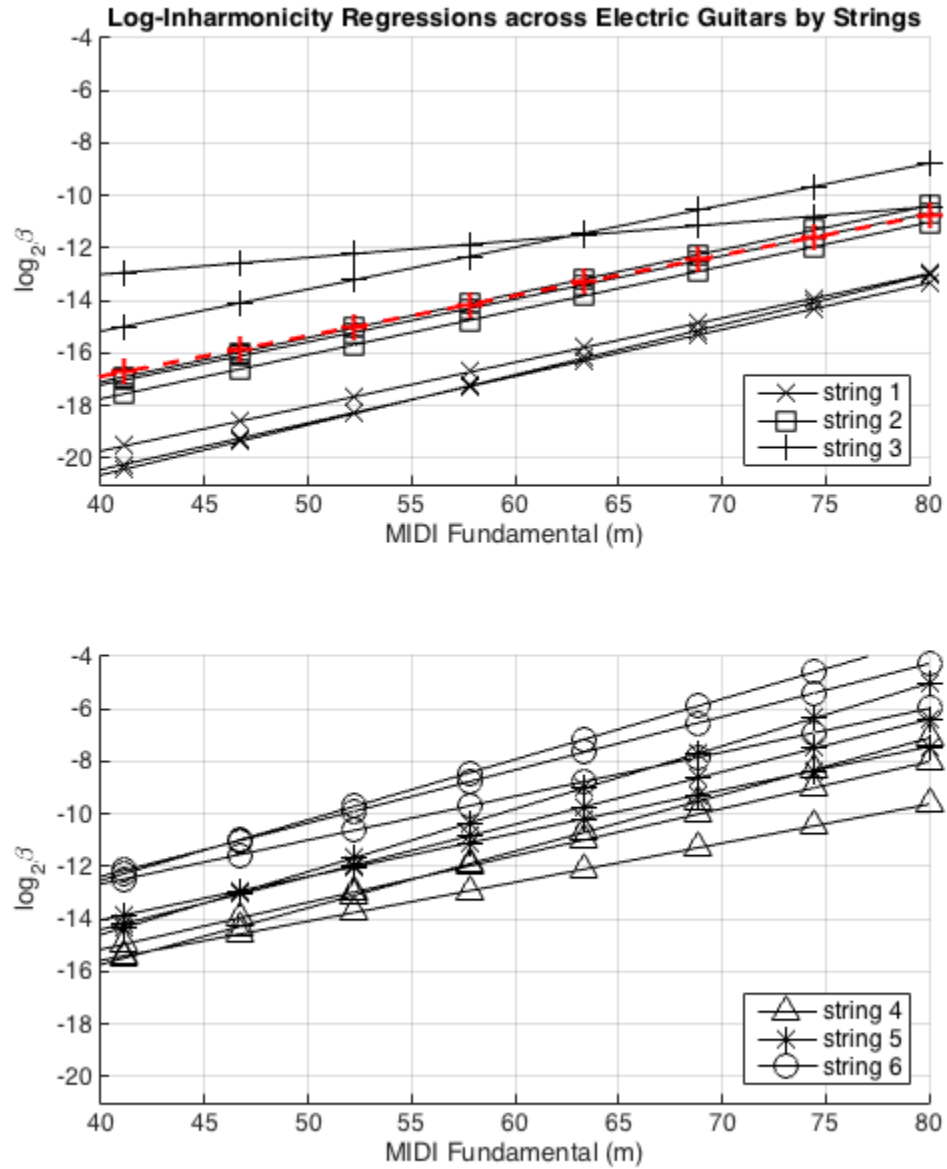
Figure 5.3: Superimposed string-wise regressions of all three RWC electric guitars. Top: strings 1, 2, and 3. Bottom: strings 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the *inconsistent* intercept and slope. In particular, note the overlap between one of string 3's regressions (dashed red) and string 2's regressions. Also observe the overlapping slope and intercept ranges for strings 3 and 4 across the two plots.
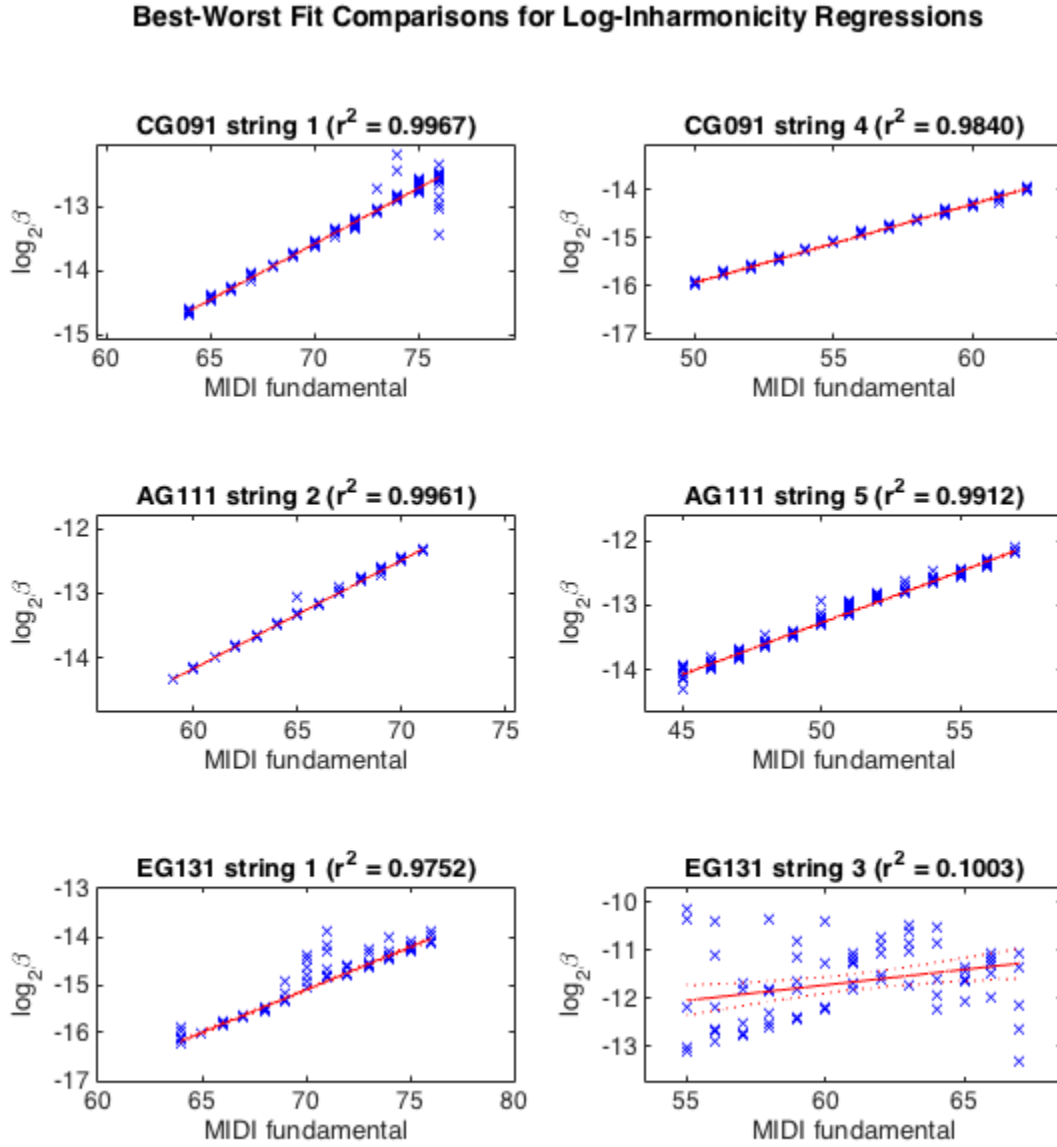
Figure 5.4: Best and worst bisquare log-inharmonicity regression fits, as measured by r^2, for representative guitars from each of the three classes. Top row: classical. Middle row: acoustic. Bottom row: electric. Left and right columns are best and worst fits, respectively.

of pickup transduction and inharmonicity estimation variability. For each guitar, and for each string, we collected note performances by common dynamic levels (piano, mezzo, and forte) and regressed their log-inharmonicities against their fundamentals, as usual. We then plotted statistics of the log-inharmonicities' residuals against their played performance intensity. Figure 5.5 illustrates these statistics for EG131; other electric guitars' boxplots were omitted because their statistics were similar. We expected that variability would increase with dynamics, but the data didn't support this hypothesis. Interquartile ranges didn't exhibit any clear trend with respect to performance intensity, nor was the magnitude of their variation noteworthy. The cause of the inharmonicity unreliability is therefore probably linked to some spectral characteristic of the electric guitars, e.g. considerably more non-tonal peaks which the MAT algorithm erroneously selects as partials, or some aspect of the pickup's frequency response that's throwing off the MAT algorithm, etc.

5.2   Limitations

It's important to highlight factors that potentially limit the significance of our claims. Our system currently operates only on clean (i.e. no effects processing), isolated, monophonic guitar recordings, which are obviously not representative of the majority of guitar audio. Multiple layers of effects processing are commonly used by performers of the instrument to achieve distinctive, creative sounds. Guitarists often play in concert with other musicians, producing complex soundscapes instead of being featured in isolation. Moreover, guitar passages are rarely in their entirety strictly monophonic. These are large obstacles that separate our RWC subset from actual guitar performances, and thus the implications of our transcription results should be considerably tempered. Noteworthy though, is the extent to which other teams have overcome some of these factors. Certain tablature transcription systems [1, 4, 9, 19] have implemented polyphonic capabilities into their systems with good success, so polyphony is perhaps a smaller barrier than the others. As part of a multipitch estimation system in [31], Yazawa deduced tablature from
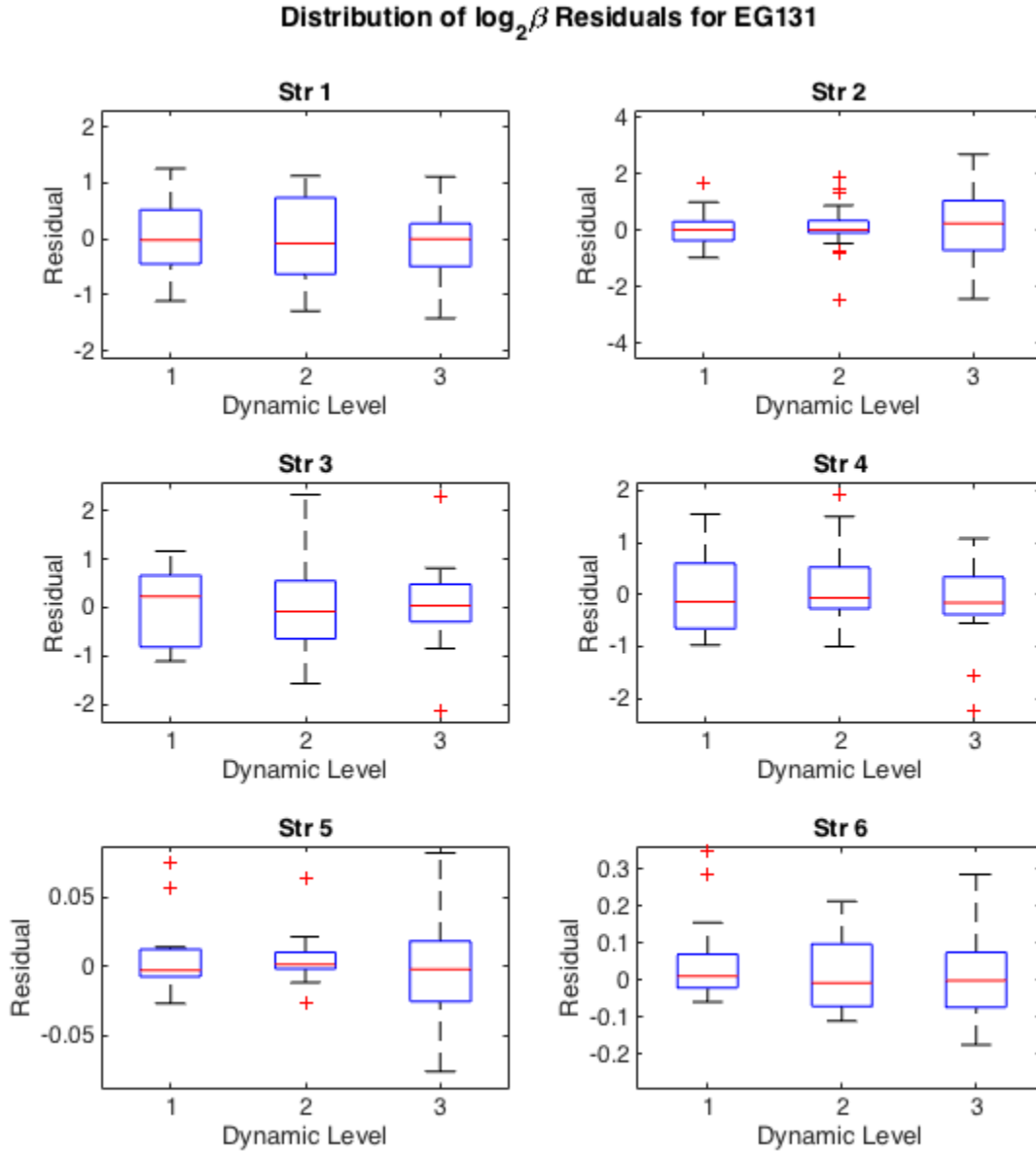
Figure 5.5: Boxplots of log-inharmonicity regression residuals for EG131 by dynamic level. Group 1: piano, Group 2: mezzo, Group 3: forte. Red center horizontal lines are medians; top and bottom edges of blue boxes represent upper and lower quartiles; whiskers depict full range of data excluding outliers; red plus signs denote outliers. Vertical axes not constant throughout plots; we're highlighting relative distributions of the boxplots here.

RWC recordings that featured guitar passages playing over backing tracks, though the tablature accuracy is not reported.

Additionally, we assumed perfect fundamental pitch estimation in this work. Our decision to simply use pitch labels of the RWC recordings stemmed from convenience and scope; the focus of this work was simply evaluation of a specific approach to examining notes' inharmonicities for tablature transcription, so not concerning ourselves with the reliability of a pitch detection routine was a luxury that expedited this thesis. That being said, it should be acknowledged that the reliability of one's pitch estimator is a sure factor in the translation of this work's results to the real world.

Furthermore, the MAT inharmonicity estimation routine we used was different from Barbancho's. The credibility of the comparison between our results and theirs is thus somewhat compromised. It's quite possible that any performance gains observed in our implementation are due in part to potentially superior inharmonicity extraction. This is, unfortunately, a factor we didn't control for, and if time had allowed we should have replicated Barbancho's system and substituted their inharmonicity estimation step with MAT to produce results that were as comparable as possible.

5.3   Future Work

A potential area for run-time performance enhancement for tablature transcription systems is the use of musical context. Musical passages on the guitar tend to exhibit continuity; it's usually the case that a riff is played in a localized area of the fretboard, for example, instead of jumping around the length of the guitar neck. Transcription systems could apply these performance-borne simplifications to their output to potentially improve accuracy, say, when a stray note is transcribed a plausible yet questionable twelve frets away from the rest of those in the riff.

A measure of objective confidence in the transcription is required to do this. Without it, there'd be no way to determine whether it was the stray note or the majority of the riff which was

in fact correctly transcribed. We investigated the amenability of our log-inharmonicity residual-minimizing classification scheme to reliable confidence measurements, and found encouraging results. Figure 5.6 displays the probability, for each of the nine RWC guitars, of our system producing a correct fretboard transcription given increasingly wide windows in which we consider a note's inharmonicity estimate to have fallen. The inverse trend between residual window width and correctness probability shows that we can reliably infer some degree of confidence in a note's transcription given its distance from its classified line. This can be leveraged in a post-processing stage to further refine tablature with isolated erroneous, yet plausible, transcriptions, and should be implemented in future systems.

Another dimension we didn't consider was varying the regression polynomial. It would be interesting to see how differently, if at all, the system would perform if the log inharmonicity transformation was forgone and an exponential fit was explored instead. This should also be investigated in future inharmonicity-based transcription research.
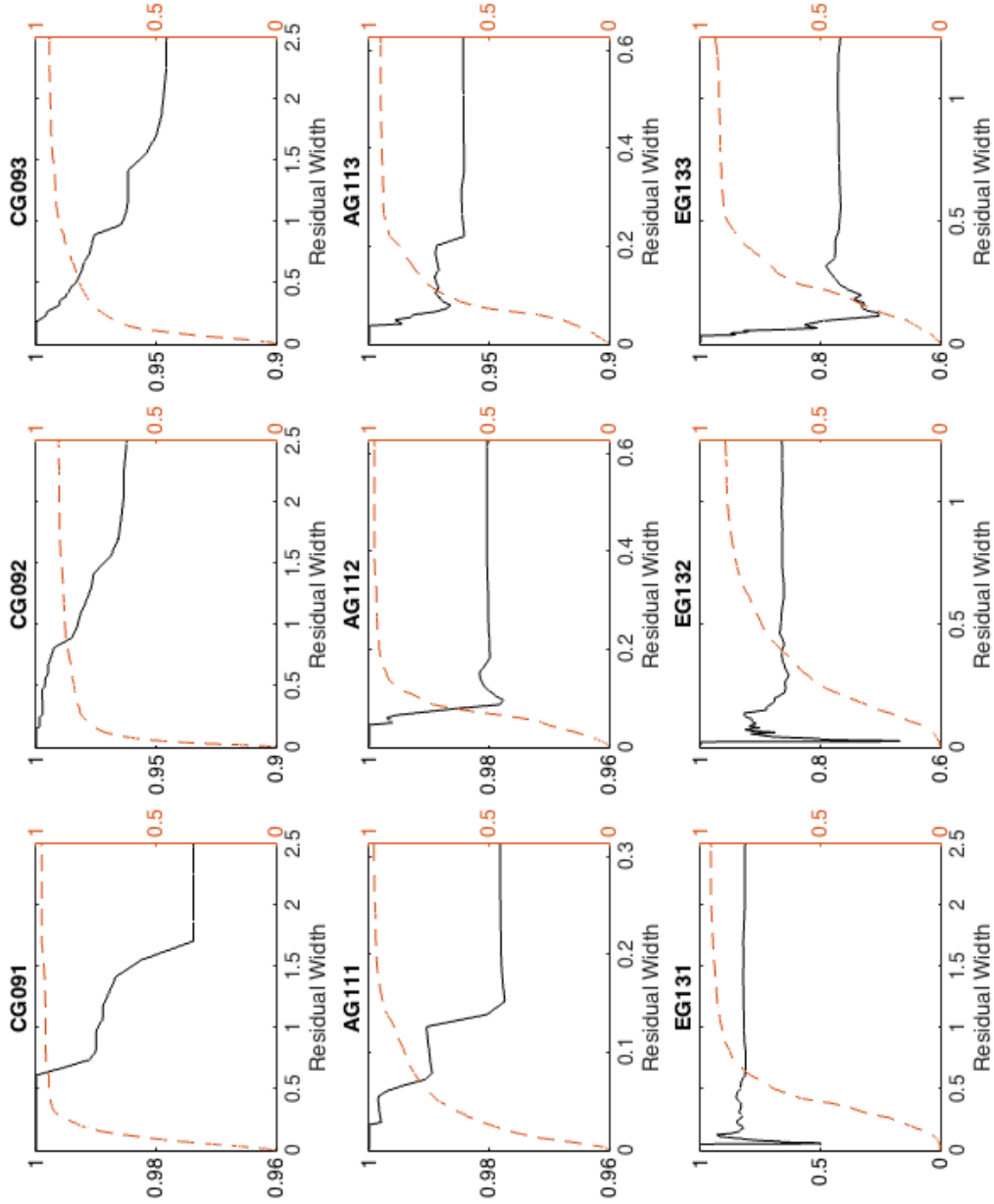
Figure 5.6: Solid black lines and left vertical axis: probability of correctly transcribing a note given the residual width under which the note's log-inharmonicity falls. Dashed orange lines and right vertical axis: CDF of the log-inharmonicity residuals, so one can get a sense of the proportion of data represented by ticks in the horizontal axis.

Chapter 6:   Summary

We began by introducing the guitar and an essential component of its musicality: the overlapping pitch ranges of its strings. We related this to tablature and its ability to uniquely specify fretboard positions for musical passages, then highlighted its popularity and its tedious manual annotation process. This led us to introduce automatic transcription systems and our goals for this work as an extension to the topic, which we reviewed in the following chapter.

We paid special attention to the partial coincidence tally (PCT) transcription method introduced by Barbancho in [4], which exploited only one feature known as inharmonicity in development of a successful system using only audio. We explained how inharmonicity affects the degree of upward skew of a note's partials, then discussed the intuition behind its discriminative power before introducing our novel approach.

After reiterating one of Barbancho's key realizations – that inharmonicity along a given string was deterministic – we showed that the logarithm of these trajectories was linear with respect to MIDI pitch. We consequently proposed characterizing guitar strings as their log-inharmonicity trajectories and performing classification (and therefore implicitly tablature transcription) by assigning to unknown notes the strings whose lines best approximated the notes' log-inharmonicities. Lastly, we derived the effect of a tuning change on these lines, and arrived at a general tuning-compensation feature applicable to any inharmonicity-based system.

Next, we reported results from experiments on a subset of the RWC music instruments database and a couple personal guitar recordings. Benchmark comparisons against the PCT method using theoretical and empirical inharmonicity trajectories revealed that our method out-

51

performed the state-of-the-art for acoustics and electrics when trained on guitars that included others in addition to the test guitar. Residual-minimization on inharmonicity trajectories therefore leverages the inharmonicity feature in a more generalizable fashion. We also validated our inharmonicity compensation feature, reporting accuracy increases on personal recordings of alternately-tuned electric and acoustic guitars.

A cursory investigation of the cause for the poorer electric guitar transcription performance revealed that inharmonicity estimation for the instrument is less reliable, and the electric guitars' pickups aren't to blame; it must rather be spectral nuances that confuse the inharmonicity estimation algorithm. We closed with limitations and future scope, cautioning against misinformed application of our reported results to real-world guitar scenarios, and highlighting musical context exploitation and different regression polynomials as potential areas for elaboration.

In conclusion, characterizing guitar strings as log-inharmonicity trajectories against pitch and classifying notes by minimizing their trajectory residuals is an effective approach to tablature transcription. Most notably, it outperforms the benchmark inharmonicity-based method on most of the RWC guitar classes when trained on reference guitars that include others in addition to the test guitar. Furthermore, adjusting inharmonicity estimates with the compensation factor introduced in this work is effective for improving transcription accuracy on alternately-tuned guitars.

Bibliography

[1] J. Abeßer. Automatic string detection for bass guitar and electric guitar. In *From Sounds to Music and Emotions - 9th International Symposium, CMMR 2012*, pages 333–352, London, UK, June 2012.

[2] A. Barbancho, A. Klapuri, L. Tardón, and Isabel Barbancho. Automatic transcription of guitar chords and fingering from audio. In *Proceedings of the IEEE Transactions on Speech and Language Processing*, pages 915–921, March 2012.

[3] I. Barbancho, A. Barbancho, L. Tardón, and S. Sammartino. Pitch and played string estimation in classic and acoustic guitars. In *Proceedings of the 126th Audio Engineering Society (AES) Convention*, Munich, Germany, 2009.

[4] I. Barbancho, L Tardón, S. Sammartino, and A. Barbancho. Inharmonicity-based method for the automatic generation of guitar tablature. In *Proceedings of the IEEE Transactions on Audio, Speech, and Language Processing*, pages 1857–1868, August 2012.

[5] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13 (5):1035–1047, 2005.

[6] G. Burlet and I. Fujinaga. Robotaba guitar tablature transcription framework. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, November 2013.

[7] Gregory D. Burlet. Guitar tablature transcription using a deep belief network. Master's

thesis, University of Alberta, Canada, 2015.

[8] A. M. Burns and M. M. Wanderley. Visual methods for the retrieval of guitarist fingering. In *Proc. Conf. New Interfaces for Musical Expression*, pages 196–199, France, June 2006.

[9] C. Dittmar, A. Mäanchen, and J. Abeßer. Real-time guitar string detection for music education software. In *Proceedings of the 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 1–4, 2013.

[10] S. Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, September 2006.

[11] H. Fletcher, E. Blackham, and R. Stratton. Quality of piano tones. *Journal of the Acoustical Society of America*, 34(6):749–761, 1962.

[12] T. Gagnon, S. Larouche, and R. Lefebvre. A neural network approach for preclassification in musical chords recognition. In *Conf. Rec. 37th Asilomar Conf. Signals, Systems, Comput.*, volume 2, pages 2106–2109, 2003.

[13] A. Galembo. Method for objective evaluation of the sound quality in piano trebles (in russian). In *Proc. of 18th Acoust. Conf. of Czechoslovakia*, 1979.

[14] A. Galembo. *The quality of piano tones (in Russian)*. Legkaya Industria Press, Moscow, 1987.

[15] A. Galembo and A. Askenfelt. Measuring inharmonicity through pitch extraction. *STL-QPSR*, 35(1):135–144, 1994.

[16] Alexander Galembo and Anders Askenfelt. *IEEE Transactions on Speech and Audio Processing*, 7(2), March 1999.

[17] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pages 229–230, October 2003.

[18] M. Hodgkinson, J. Wang, J. Timoney, and V. Lazzarini. Handling inharmonic series with

median-adjustive trajectories. In *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*, Como, Italy, September 2009.

[19] C. Kehling, J. Abeßer, C. Dittmar, and G. Schuller. Automatic tablature transcription of electric guitar recordings by estimation of score- and instrument-related parameters. In *Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, September 2014.

[20] C. Kerdvibulvech and H. Saito. Vision-based detection of guitar players fingertips without markers. In *Comput. Graph., Imag., Visualisat. (CGIV)*, pages 419–428, 2007.

[21] R. Macrae and S. Dixon. A guitar tablature score follower. In *Proc. Int. Conf. Multimedia Expo*, pages 725–726, 2010.

[22] P. O'Grady and S. Rickard. Automatic hexaphonic guitar transcription using non-negative constraints. In *Proc. IET Irish Signals Syst. Conf. (ISSC 2009)*, pages 1–6, Dublin, Ireland.

[23] M. Paleari, B. Huet, A. Schutz, and D. Slock. A multimodal approach to music transcription. In *Proc. 15th IEEE Int. Conf. Image Process (ICIP)*, pages 93–96, 2008.

[24] D. Radicioni, C. D. Scienza, and V. Lombardo. Guitar fingering for music performance. In *Proc. Int. Comput. Music Assoc.*, pages 527–530, 2005.

[25] A. Radisavljevic and P. Driessen. Path difference learning for guitar fingering problem. In *Proc. Int. Comput. Music Conf. (ICMC04)*, 2004.

[26] J. Rauhala, H. M. Lehtonen, and V. Välimäki. Fast automatic inharmonicity estimation algorithm. *Journal of the Acoustical Society of America*, 121(5):184–189, 2007.

[27] S. Sayegh. Fingering for string instruments with the optimum path paradigm. *Computer Music Journal*, 13(3):76–84, 1989.

[28] K. M. Short and R. A. Garcia. Signal analysis using the complex spectral phase evolution (CSPE) method. In *Audio Engineering Society 120th Convention*, Paris, France, May 2006.

[29] The MathWorks. Least-Squares Fitting Documentation. http://www.mathworks.

`com/help/curvefit/least-squares-fitting.html`, June 2017.

[30] D. Tuohy and W. D. Potter. Generating guitar tablature with LHF notation via DGA and ANN. In *Advances in Applied Artificial Intelligence: 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2006, Annecy, France, June 27-30, 2006, Proceedings*, pages 244–253, 2006.

[31] K. Yazawa, D. Sakaue, K. Nagira, K. Itoyama, and H. Okuno. Audio-based guitar tablature transcription using multipitch analysis and playability constraints. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 196–200, 2013.